universitätbonn
Rheinische
Friedrich-Wilhelms-
Universität Bonn

LIMES
Life & Medical Sciences Institute

Transcriptional characterization of macrophages reveals unexpected novel biology

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Jia Xue

aus

Jiangsu, China

Bonn

Oktober, 2015

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Joachim L. Schultze

2. Gutachter: Prof. Dr. Martin Hofmann-Apituis

Tag der Promotion: 27.10.2015

Erscheinungsjahr: 2015

Preface

This dissertation is completely based on my personal work and large national and international collaborations. It consists of four related topics of biological studies on macrophages using bioinformatic methods.

Chapter 4 attempts to explore human macrophage activation systematically via large transcriptomic data from diverse signal inputs using systems biology methods. Parts of **Chapter 4** were published by me as first author in Immunity (Xue et al., 2014) together with some additional work, which is not included in this thesis. For this publication I was responsible for all major areas including concept formulation, data collection and analysis, as well as manuscript composition.

Chapter 5 is an extensive study on the same subject not yet published. For example this study refines the common regulatory networks identified in **Chapter 4** by integrating multiple inference methods. The same human macrophage dataset was used. The results of this study are currently prepared as a manuscript for submission. Parts of the study have been presented as an oral presentation as well as a poster at the Keystone Symposium "Dendritic Cells and Macrophages Reunited" in Montreal.

Chapter 6 summarizes transcriptome-based network analysis of human peripheral blood mononuclear cells (PBMC) stimulated with bacterial and fungal pathogens. This is an international collaboration project with the group of Prof. Mihai G. Netea from the Department of Medicine, Radboud University Nijmegen Medical Centre, The Netherlands. I did the bioinformatic analysis of the large microarray data and we are currently preparing a manuscript to publish the results.

Chapter 7 extends from human *in vitro* work to murine *in vivo* data. It is also based on a large dataset established in collaboration with the group of Prof. Dr. Marco Prinz at the Institute of Neuropathology, University Hospitals of Freiburg. For this project, my task was to analyze the transcriptional dataset to describe the role of the transcription factor Irf8 in myeloid cells during embryogenesis and in adult tissue cells and to identify novel markers for further experimental validation. The results will be prepared as a manuscript for publication as well.

I declare that the dissertation is written completely by myself. None of the text is taken directly from previously published or collaborative articles.

Contents

Summary

Macrophages are very plastic and versatile immune cells in response to different environmental signals. Similar phenomenon has been observed for other myeloid compartment, such as monocytes. Moreover, it has been described that different tissue macrophage subpopulations have distinct origins. In this dissertation, I systematically analyzed a large resource dataset to assess transcriptional regulation during human macrophage activation by comparing a diverse set of stimuli on a single microarray platform under highly standardized conditions. Network modeling of this dataset led to the extension of the current M1 versus M2 polarization model to a “multi-dimensional model” with at least nine distinct macrophage activation programs. Applying these transcriptional programs to human *in vivo* alveolar macrophages from smokers and patients with chronic obstructive pulmonary disease (COPD) revealed an unexpected biology. Reverse engineering of large transcriptional dataset by integrating multiple network inference approaches sharpens the resolution of the common macrophage activation regulatory networks. And the refined network indicated that transcription factors are the most important components in regulatory circuits involved in macrophage activation. Furthermore, by applying the same computational methodologies to a transcriptomic dataset of infected human peripheral blood mononuclear cells (PBMC), I extended my studies to identify common and stimulus-specific transcriptional programs in host defense against bacteria and fungi. By combination of knowledge-based and data-driven analysis, I propose refined pathway models for these microbial infections on transcriptional level. Finally, computational studies on gene expression profiles for embryonic and adult tissue macrophages from both wild type and Irf8-deficient mice revealed distinct origins and transcriptional profiles of different tissue macrophage subpopulations and a crucial role of Irf8 in macrophage ontogeny and homeostasis. Overall, applying systems biology approaches, especially advanced methods on large enough transcriptional datasets enables robust and accurate statistical predictions. Thus, the studies on macrophages or myeloid cells using these approaches successfully uncovered the complex dynamic regulatory networks of these cells and reflected a hitherto unexplored biology.

1 Introduction

Macrophage biology has been attracted an enormous interest during the last 5 years – in part due to the recognition that these cells are not only responsible for phagocytosis but fulfill a myriad of other functions. Furthermore, thanks to the rapid development of high-throughput data generation techniques and corresponding data analytic tools during last two decades, genome-wide assessments of genetic, epigenetic and transcriptional changes allow to better understand macrophage biology.

There are numerous open questions in macrophage and myeloid cell biology ranging from aspects of their origin, their differentiation, to their activation. In three independent studies, by analyzing three independent transcriptional profiling datasets, I have addressed the questions: how human macrophage activation can be better described by global transcriptional profiling, what the role of myeloid cells like monocytes and macrophages play in the response to infectious pathogens and in a murine *in vivo* system. Moreover, I have studied the role of a particular transcription factor Irf8 during macrophage development and in different adult tissue macrophage populations.

Macrophages are versatile immune cells that are involved in inflammation, diseases and maintenance of tissue homeostasis. In the last two decades, a conceptual paradigm for the characterization of macrophage activation, i.e. M1 (classical activation) versus M2 (alternative activation) polarization model representing two extreme states of signals collected by macrophages, has been established (Biswas and Mantovani, 2010). The dichotomous system has been very helpful in describing immune responses during acute infections, allergies, asthma, and obesity (Chinetti-Gbaguidi and Staels, 2011). However, more and more observations from macrophages involved in chronic inflammation, chronic infection, cancer and other diseases strongly suggest that the myeloid compartment has a much broader and more complex transcriptional repertoire relying on the different microenvironmental signals they obtain (Boorsma et al., 2013; Chow et al., 2011; Edin et al., 2012; Hodge et al., 2011; Lawrence and Natoli, 2011; Martinez et al., 2009; Mosser and Edwards, 2008; Murray and Wynn, 2011; Noy and Pollard, 2014; Ostuni et al., 2015; Reinartz et al., 2013; Vlahos and Bozinovski, 2014). In spite of many genomic studies analyzing macrophage activation in response to toll-like receptor (TLR) ligands, microbial pathogens and cytokines, heretofore nearly no attempts have been made to reconcile these

observations by building new and integrative models of macrophage activation (Martinez et al., 2006; McDermott et al., 2011; Nau et al., 2002; Ramsey et al., 2008).

The immune system plays crucial roles in bacterial and fungal infections. Peripheral blood mononuclear cells (PBMC) are the major components, which quickly give responses to these pathogens. These responses on the one hand undergo stimulus-specific mechanisms; however, on the other hand, share particular intracellular signaling pathways, which ultimately lead to host protective effects such as production of pro-inflammatory cytokines, antigen processing and presentation.

Other than inflammatory conditions, the origins of different tissue macrophages and their embryonic development are also an important aspect to study macrophage biology. It is known that tissue-resident macrophages stand at the frontline of tissue defense where they are discretely distributed and transcriptionally programmed to encounter with environmental signals. Different subsets of tissue macrophages originate from different origins, i.e. primitive hematopoiesis and definitive hematopoiesis (Ginhoux and Jung, 2014). And these processes are controlled by a few global regulators such as Pu.1, Myb and Irf8 (Kierdorf et al., 2013).

The aim of this PhD study is to precisely characterize human macrophage activation from diverse signal inputs and to refine the pathway models of host defense against various pathogens on transcriptomic level. With the mouse data I attempt to transcriptionally illustrate the distinct tissue macrophage populations and the role of one transcription factor Irf8 in macrophage development and adult tissue macrophages in mice.

2 Background

2.1 Blood cells and mononuclear phagocyte system

Blood cells (hematocytes), which are produced by hematopoiesis, are classified into three general categories: red blood cells (erythrocytes), white blood cells (leukocytes) and platelets (thrombocytes). Erythrocytes normally live for about 120 days in humans and the role is to carry oxygen, collect carbon dioxide via hemoglobin, and transport them throughout the body. Additionally they are positioned around the leukocyte to protect the healthy cells. Thrombocytes are very small, irregularly shaped clear cell fragments involved in hemostasis, which results in the formation of blood clots. Leukocytes are involved both in innate and adaptive immune responses to defend the body against both infectious diseases and foreign materials (Tortora and Derrickson, 2012).

By the late 1960s, the mononuclear phagocyte system (MPS) was initially formulated, constituting monocytes and macrophages (Vanfurth and Cohn, 1968). Till the discovery of the dendritic cell (DC) by Steinman in the 1970s, monocytes, macrophages and DCs have been grouped together although they have distinct morphology, function and origin (Vanfurth et al., 1972). However, lineage-tracing studies demonstrated that under homeostatic conditions, most macrophages in adults are not reliable on blood monocytes but depend almost exclusively on self-renewal (Hashimoto et al., 2013; Yona et al., 2013). Moreover, classical DCs originate from adult hematopoietic stem cells (HSC)-derived DC precursors, which differ from classical monocytes (Naik et al., 2013; Schraml et al., 2013). Therefore, not as originally proposed, instead, MPS encompasses three main groups of cells: CDP-derived DCs, embryonic-derived macrophages and monocyte-derived cells (**Figure 2.1**).

The nomenclature of MPS has been challenged for a long time. Recently, Guilliams *et al.* proposed a two-level nomenclature system, which is, mononuclear phagocytes should be first defined based on their ontogeny and then be classified on the basis of their function, location and phenotype (**Figure 2.1**) (Guilliams et al., 2014).

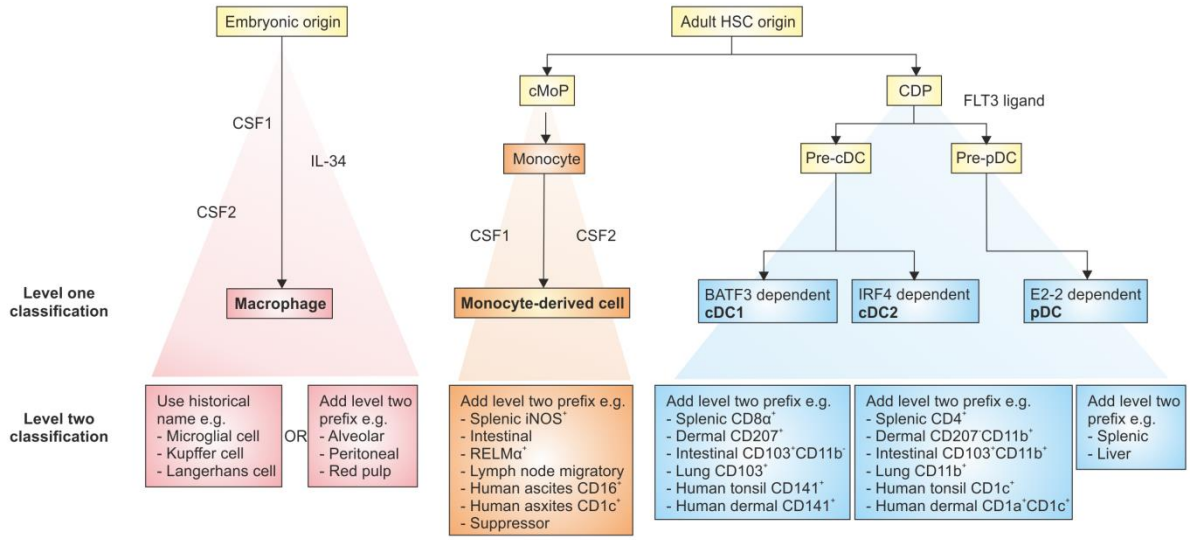

Figure 2.1 Two-level nomenclature system to classify mononuclear phagocytes. Figure was adapted from (Guilliams et al., 2014). It was suggested by the authors that mononuclear phagocytes should first be defined on the basis of their ontogeny (level one nomenclature; yellow boxes), followed by their function, location and/or morphology (level two nomenclature; pink, orange and blue boxes). Due to distinct dominant transcription factors during development and discrete committed precursors, DCs could be further subdivided into classical type 1 DCs (cDC1s, BATF3 dependent), cDC2s (IRF4 dependent) and plasmacytoid DCs (pDCs, E2-2 dependent). BATF3: basic leucine zipper transcriptional factor ATF-like 3; cMoP: common monocyte progenitor; CSF1: colony-stimulating factor 1 (also known as M-CSF); CSF2: colony-stimulating factor 2 (also known as GM-CSF); FLT3: FMS-like tyrosine kinase 3; HSC: hematopoietic stem cell; IL-34: interleukin-34; iNOS: inducible nitric oxide synthase; IRF4: interferon-regulatory factor 4; RELMα: resistin-like molecule-α.

2.2 Peripheral blood mononuclear cells and monocytes

Peripheral blood is the flowing, circulating blood of the body. Peripheral blood mononuclear cells (PBMC) are blood cells containing a rounded nucleus, including T cells, B cells, natural killer (NK) cells, monocytes, macrophages and DCs, with a few contaminants of neutrophils and mast cells. They are central players in the immune system to fight infections and invaders. Myeloid compartments such as monocytes participate in tissue healing, clearance of pathogens and dead cells, and initiation of adaptive immunity. However, recruited monocytes can also contribute to the pathogenesis of infection and chronic inflammatory disease.

Monocyte subsets were discovered by means of flow cytometry with CD14 and CD16 antibodies in human blood more than 20 years ago and their phenotype and function has

been characterized in detail in health and disease. The first evidence for heterogeneity of monocytes was the discovery of $CD16^+$ monocytes provided in the late 1980s (Passlick et al., 1989). The $CD16^-$ classical monocytes ($CD14^+$) cover 90% of the major population and the $CD16^+$ cells account for about 10% of all monocytes under physiological conditions. The number of the $CD16^+$ monocytes can vary strongly under various conditions (Ziegler-Heitbrock, 2014). For many inflammatory conditions an increase to more than 50% of all monocytes for severe infection has been observed (Fingerle et al., 1993). A recent nomenclature proposal has further subdivided the $CD16^+$ monocytes into non-classical and intermediate cells ($CD14^+CD16^+$ population) (Ziegler-Heitbrock et al., 2010).

2.3 Macrophages as part of the innate immune system

Macrophages were first discovered in the late 19th century by Ilya Metchnikoff and are evolutionary conserved phagocytes that evolved more than 500 million years ago. Macrophages are distributed in tissues throughout the body and contribute to both homeostasis and disease.

2.3.1 Monocyte-derived macrophages

It has long been known that monocytes are a systemic reservoir of myeloid precursors for generation of macrophages or DCs in response to various stimuli. But only until the middle of the 1990s people discovered colony stimulating factor-1 (CSF-1, also known as macrophage colony stimulating factor [M-CSF]) as a key factor for differentiating monocytes into macrophages in mice (Cecchini et al., 1994; Wiktor-Jedrzejczak and Gordon, 1996), *in vitro* generation of macrophages from blood monocytes became possible for mammalian cells. Another CSF family molecule, CSF-2, or granulocyte-macrophage colony stimulating factor (GM-CSF), was considered to control the development and homeostasis of the macrophage-DC lineage but to be dispensable for monocyte development (Auffray et al., 2009). However, it is also used for monocyte-macrophage/DC differentiation.

2.3.2 Tissue-resident macrophages

Most tissues in the body encompass tissue-resident macrophages, which are extremely heterogeneous populations because they fulfill tissue-specific and microenvironment-specific functions during development and adulthood. It is widely known that tissue-resident macrophages play an important role as immune sentinels in the frontline of tissue defense

where they are discretely distributed and transcriptionally programmed to encounter with pathogens or environmental stresses (Davies et al., 2013). Recent research has shown that other than blood macrophages, most adult tissue macrophages originate during embryonic development but not from circulating monocytes. The composition and functions of embryonically derived and adult-derived macrophages are dependent on different tissues (**Table 2.3**). Nevertheless, it still needs to be find out whether macrophages from distinct origins have interchangeable functions or if they have unique effects at steady state (Epelman et al., 2014b). Tissue macrophages modulate homeostasis and normal physiology through their regulation of diverse activities, including metabolism and neural connectivity, and by detecting damage. However, the normal physiology of tissue macrophages can be disrupted by continuous insult in many cases and thereby give rise to macrophages associated diseases, such as cancer, metabolic disorders and obesity (Sagaert et al., 2012; Wynn et al., 2013).

The multiple waves of hematopoiesis during development are classified into two stages (**Figure 2.3.1**). The initial wave of blood production in the mammalian yolk sac is called primitive hematopoiesis. Primitive hematopoietic progenitors are generated in the yolk sac blood islands and produce primitive (yolk sac-derived) macrophages from embryonic day 8.5-9.0 (E8.5-E9.0). These primitive hematopoietic progenitors also spread into embryos to establish the blood circulatory system and colonize the whole embryo from E9.0-E10.0 to generate fetal primitive macrophages, which is dependent on the transcription factor PU.1 but independent of Myb. The primitive hematopoietic system is transient and rapidly replaced by adult-type definitive hematopoiesis. The wave of definitive hematopoiesis occurs in the aorta–gonads–mesonephros (AGM) from E10.5 and is responsible for the generation of hematopoietic stem cells (HSCs) with multi-lineage hematopoietic potential, which is responsible for formation of the fetal liver. This process is dependent on Myb (Orkin and Zon, 2008). Around E11.5-E12.5, hematopoiesis in the fetal liver generates monocytes, which start to migrate into the blood from E12.5-E13.5 and then attack embryonic tissues around E13.5-E14.5. Nonetheless, yolk sac progenitors can also contribute to fetal liver hematopoiesis during the middle phase of embryonic development (from E8.5) and might be involved in the generation of fetal liver monocytes. Once inside the tissues, in presence of colony-stimulating factor 1 receptor (CSF1R), fetal liver-derived monocytes proliferate and differentiate into macrophages and further become sub-populations of adult tissue

macrophages such as alveolar macrophages in the lung (Guilliams et al., 2013) and cardiac macrophages in the heart (Epelman et al., 2014a). However, microglia arise predominantly from yolk sac-derived macrophages (Ginhoux et al., 2010); and Langerhans cells have two types of origins, i.e., they mostly originate from fetal liver-derived monocytes but also keep a few yolk sac-derived ones (Hoeffel et al., 2012). These findings suggest that the origin of adult macrophages in the steady state can be divergent considerably between tissues and tissues like the skin and gut contain adult monocyte-derived macrophages.

Microglia, parenchymal tissue-resident macrophages in the central nervous system, are important mediators in tissue homeostasis in health and disease. Microglia are sustained and self-renew within host tissues throughout adulthood independent of progenitors from the bone marrow. In yolk sac from E8.0-9.0, erythromyeloid progenitors (EMP) are developed into immature yolk sac macrophages (A2) via early myeloid progenitors (A1) dependently of PU.1 and Irf8. Once mature microglia are formed (E10.5), matrix metallopeptidases MMP8/9 become key regulators in phagocytosis and migration. During maturation in the embryo, the proliferative capacity of microglia is decreasing (**Figure 2.3.2**) (Kierdorf et al., 2013).

Table 2.3 Macrophages in tissues, homeostasis and disease

Tissue	Tissue macrophage populations: origins in mice	Normal physiology	Pathology
Brain	Microglia: Yolk sac	Neuronal patterning, fluid balance	Neurodegeneration
Bone marrow	Osteoclasts and macrophages: Bone marrow	Bone remodeling, hematopoiesis	Osteoporosis and osteopetrosis, leukemia
Lung	Alveolar macrophages: HSC (embryonic) $CD11b^+$ macrophages: unknown	Immunity	Asthma, COPD
Liver	Kupffer cells: Yolk sac + HSC (embryonic) $CD11b^+$ macrophages: unknown	Lipid metabolism, toxin removal	Fibrosis
Heart and vasculature	$CCR2^-$ macrophages: Yolk sac + HSC (embryonic and adult) $CCR2^+$ macrophages: HSC (adult)	Angiogenesis	Atherosclerosis
Spleen	Red pulp macrophages: HSC (embryonic) Marginal zone macrophages: unknown	Branching morphogenesis	Cancer and metastasis
Kidney	Kidney macrophages: HSC (embryonic and adult?)	Immunity	Arthrotos, EAE, IBD
Intestine	Intestinal macrophages: HSC (adult)	Maintaining mucosal homeostasis	Infectious diseases, metabolic disorders, inflammatory bowel diseases, malakoplakia
Skin	Langerhans cells (LC): Yolk sac + HSC (embryonic) Dermal macrophages: HSC (adult)	Antigen processing and presentation	LCH
Peritoneum	Peritoneal macrophages: HSC (embryonic)	Immunity	Peritoneal infections
Adipose tissue	Adipose tissue macrophages: Bone marrow	Metabolism, adipogenesis	Obesity and diabetes

HSC, hematopoietic stem cell; EAE, experimental autoimmune encephalomyelitis; IBD, inflammatory bowel disease; LCH: Langerhans cell histiocytosis.

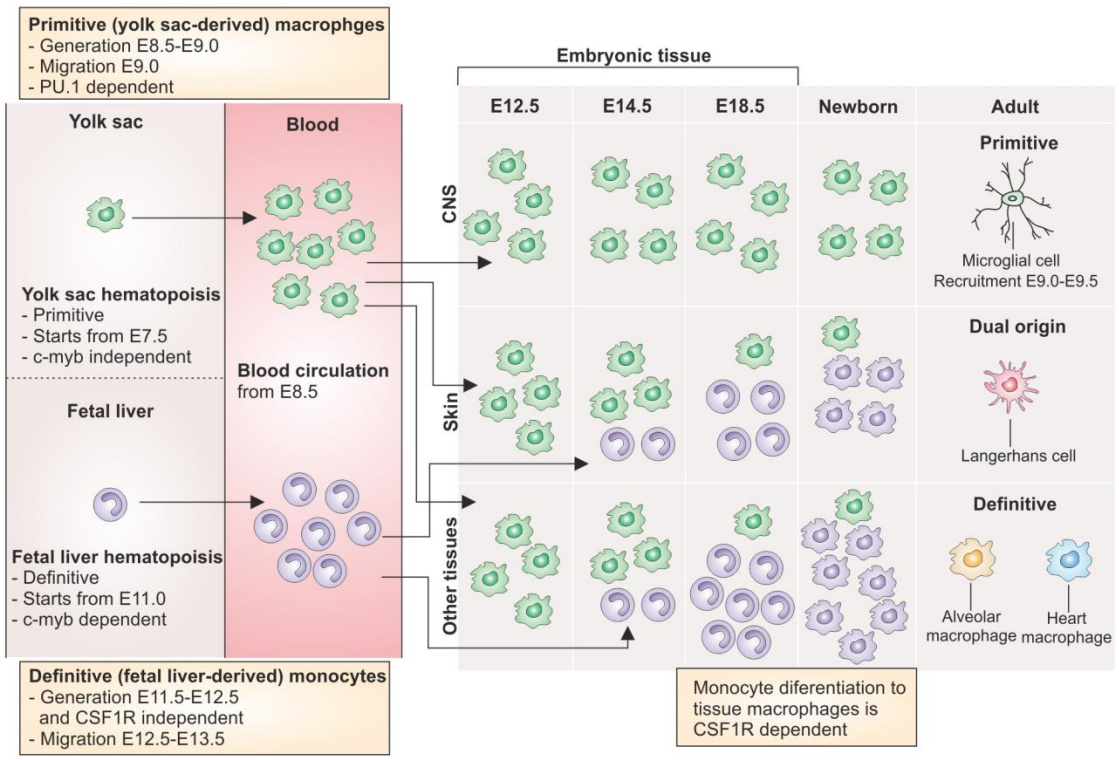

Figure 2.3.1 Embryonic macrophage development in mouse. Figure was adapted from (Ginhoux and Jung, 2014). CNS: central nervous system; CSF1R: colony-stimulating factor 1 receptor.

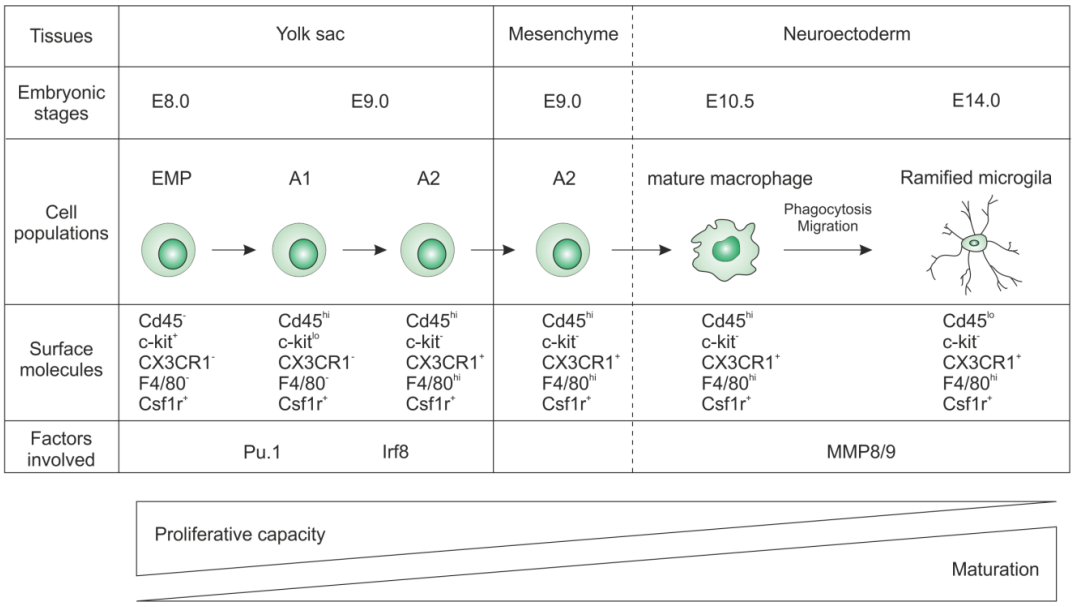

Figure 2.3.2 Microglial development in mouse. Figure was suggested by (Kierdorf et al., 2013). During maturation of microglia in the embryo, the proliferative capacity is decreasing. EMP: erythromyeloid progenitor; A1: early myeloid progenitor; A2: immature yolk sac macrophage.

2.4 Inflammation and macrophage activation

Inflammation is a protective immunovascular response to external or internal stimuli, such as pathogens, damaged cells or pain. It involves immune cells, blood vessels, and molecular mediators and is considered as a mechanism of innate immunity, which is unspecific to each pathogen. The goal of inflammation is to drop down the risk of cell injury by eliminating its initial cause, cleanup cells and tissues damaged from the original offence and inflammatory processes, and also to initiate tissue repair (Abbas and Lichtman, 2009).

Inflammation is often categorized as either acute or chronic. Acute inflammation is the initial response of the body to harmful stimuli and is achieved by the increased movement of plasma and leukocytes (especially granulocytes) from the blood into the injured tissues. The duration is only a few days. Chronic inflammation, however, often lasting for up to months or years, leads to a progressive shift in the type of cells present at the site of inflammation and is characterized by simultaneous destruction and healing of the tissue from the inflammatory process. The major involved players are fibroblasts and mononuclear cells including monocytes, macrophages, lymphocytes and plasma cells (Abbas and Lichtman, 2009).

Macrophages, as major players in inflammation, respond and behave differently when receiving different microenvironmental signals. Cytokines and microbial products profoundly affect the function of macrophages and other mononuclear phagocytes.

2.4.1 Classical and alternative activated macrophages

A decade ago, the M1 (classical activation) and M2 (alternative activation) paradigm was established as mirror states for Th1 and Th2 polarization of T cells (Biswas and Mantovani, 2010; Mantovani et al., 2002). Moreover, much progress has been made in defining the signaling pathways, transcriptional networks, and epigenetic mechanisms underlying this bipolar system.

The M1 phenotype, triggered by stimulation of interferon-γ (IFNγ), was first discovered in 1983 (Nathan et al., 1983). In the following years, studies on these macrophages revealed their characteristics such as the high expression of pro-inflammatory cytokines (e.g. tumor necrosis factor [TNF] and interleukin-1β [IL1β]), high production of reactive nitrogen and oxygen intermediates, promotion of Th1 response and strong microbicidal and tumoricidal

activity. A few years after discovery of Th1 and Th2 in T cell immunity (Mosmann and Coffman, 1989), an alternative activation of macrophages was proposed and further described as M2 macrophages (Abramson and Gallin, 1990; Stein et al., 1992). Driven by IL4/IL13 stimulation, they are described to have anti-inflammatory effects and immunoregulatory functions and induce Th2 response. They are characterized by efficient phagocytic activity, high expression of scavenging molecules, the expression of mannose and galactose receptors, production of ornithine and polyamines through the arginase pathway (Gordon and Martinez, 2010). They play essential roles in parasite control and promotion of tissue remodeling and tumor progression. M1 and M2 macrophages have distinct chemokine expression profiles (Sica and Mantovani, 2012). Additionally, cells stimulated with stimuli including IL10, glucocorticoid hormones, immune complexes and molecules released from apoptotic cells were also considered to be "M2-like" because they share some properties with M2 cells (e.g. high expression of mannose and scavenger receptor) although with different chemokine production (Biswas and Mantovani, 2010).

It is a complex repertoire of signaling molecules, transcription factors, post-transcriptional regulators and epigenetic mechanisms such as histone modifications, modulating each state of macrophage activation. Activation of canonical IRF/STAT signaling pathways by IFNγ and toll-like receptor (TLR) signaling skews macrophage function towards the M1 phenotype via transcription factor STAT1 while activation by IL4/IL13 leads to the M2 phenotype via STAT6 (Sica and Bronte, 2007).

2.4.2 Stimuli associated with chronic inflammation

Chronic inflammation, a mechanism of innate immunity, is a protective immunovascular response with longer duration. It is often caused by persistent injury or infection (e.g. tuberculosis), autoimmune diseases (e.g. rheumatoid arthritis and multiple sclerosis) and prolonged exposure to a toxic agent (e.g. pulmonary silicosis) (Kumar et al., 2013). Phagocytic cells infiltrated in chronic inflammation caused by infectious pathogens including intracellular bacteria, for instance, *Mycobacterium tuberculosis* (Flynn and Chan, 2001). Macrophages and neutrophils play a major role during early immune responses against these intracellular pathogens (Grivennikov et al., 2005; Tsai et al., 2006). Previous research showed that the combination of the factors TNF, prostaglandin E2 (PGE_2) and TLR2 ligand P3C (TPP) induced expression of genes including CD25, COX-2, IL10, and indoleamine 2,3-dioxygenase

(IDO) in macrophages in human granulomatous structures, suggesting that these host factors shape the transcriptional program during chronic inflammation (Popov et al., 2006; Popov et al., 2008).

2.4.3 Non-classical stimuli such as free fatty acids

Fatty acids are carboxylic acids with a long aliphatic chain, which is either saturated or unsaturated (**Table 2.4**). Free fatty acids (FFAs) that are not attached to other molecules are generally derived from the breakdown triglyceride molecules. They are usually released from adipose tissues and transported in the blood. Under certain conditions of metabolic dysfunction, cellular components of the innate immune system may be activated by FFAs in the absence of external pathogens, leading to pathologic consequences. For example, macrophages play a crucial role in the initiation of a chronic inflammatory state in obesity which leads to insulin resistance. Already in 2001, it has been observed in mouse models that omega-3 fatty acids are anti-inflammatory, polyunsaturated fatty acids are weak or neutral, and saturated fatty acids are pro-inflammatory (Lee et al., 2001). In response to increases in FFA release from obese adipose depots, pro-inflammatory macrophages infiltrate adipose tissues. These pro-inflammatory macrophages trigger inflammatory signaling and stress responses within cells that signal via TLR4 through JNK or IKKβ pathways, resulting in insulin resistance. If over-nutrition persists, mechanisms that counteract inflammation (such as PPAR signaling) are suppressed, and the inflammation becomes chronic (Patel et al., 2013).

Table 2.4 Examples of saturated and unsaturated fatty acids

Category	Common name	Chemical structure
Saturated fatty acids	Lauric acid (LA)	$CH_3(CH_2)_{10}COOH$
	Palmitic acid (PA)	$CH_3(CH_2)_{14}COOH$
	Stearic acid (SA)	$CH_3(CH_2)_{16}COOH$
Unsaturated fatty acids	Oleic acid (OA)	$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$
	Linoleic acid (LiA)	$CH_3(CH_2)_4CH=CHCH_2CH=CH(CH_2)_7COOH$

2.4.4 High density lipoprotein (HDL)

HDL is an 8-11 nm high-density (1.063-1.210 g/ml) lipoprotein with 40-50% protein, 25% phospholipids, 15% cholesterol and 5% triglycerides. The major protein component of HDL particles is APOA1, which functions as an acceptor for phospholipids and cholesterol on

hepatocytes, enterocytes and macrophages. HDL particles carry cholesterol from peripheral tissue to the liver. It is known that HDL has potent anti-inflammatory properties and is able to sever the links between cholesterol accumulation and inflammation, thereby being protective against atherosclerosis and other inflammatory and metabolic diseases (Tall and Yvan-Charvet, 2015). Recent studies on HDL-activated macrophages demonstrated that the protective effects of HDL against TLR-induced inflammation were fully dependent on the transcriptional regulator ATF3, rather than the quantity of HDL itself (De Nardo et al., 2014).

2.5 Host defense against microbial pathogens

Macrophages as a member of myeloid cells, however, do not work independently of their microrenvironment. Other mononuclear cells such as monocytes can also respond rapidly to foreign materials or interact with the primarily affected cells. In the next paragraphs, I will introduce previous researches on the response of peripheral blood derived human monocytes within peripheral blood derived mononuclear cells to four common microbial pathogens or pathogenic stimuli that can attack the human immune system. I will also elute to the molecular pathways that induce host protective effects by these mononuclear cells during these model infections.

Monocytes are crucial participants for an effective immune response to most pathogens. Chemokine receptors and adhesion receptors modulate monocytes migrating from the blood to the site of infection or injury (Karlmark et al., 2012). It was previously suggested that monocytes use the chemokine receptor CCR2 only for their recruitment to infected tissues. However, experiments in murine models revealed that monocytes also require CCR2 to emigrate from the bone marrow into the blood circulation (Dunay et al., 2008; Serbina and Pamer, 2006). Furthermore, circulating TLR-ligands can induce the production of MCP-1, which is the major binding partner for CCR2, by the bone marrow mesenchymal stem and progenitor cells during infection (Shi et al., 2011). These mechanisms enable monocytes to exit and migrate toward the site of infection.

2.5.1 Bacterial pathogens

Bacteria are one of the major external stimuli that affect the immune system. Many bacterial pathogens can lead to systemic infections with peripheral blood-derived immune cells being involved in host defense. For instance, LPS is one of the best studied immunostimulatory

components of bacteria, which can induce systemic inflammation and sepsis in presence of excessive signals (Beutler and Rietschel, 2003). LPS is an important structural component in the outer membrane of Gram-negative bacteria e.g. *Escherichia coli* (*E. coli*). LPS stimulation of mammalian cells requires the interactions between several proteins including the LPS binding protein (LBP), CD14, MD-2 and TLR4 (Gioannini and Weiss, 2007; Miyake, 2007). LBP is a soluble shuttle protein, which directly binds to LPS and facilitates the association between LPS and CD14 (Tobias et al., 1986; Wright et al., 1989). CD14 facilitates the transfer of LPS to the TLR4/MD-2 receptor complex and modulates LPS recognition (Wright et al., 1990) while MD-2 non-covalently associates with TLR4 but can directly form a complex with LPS in absence of TLR4 (Nagai et al., 2002). Although there is no evidence suggesting that TLR4 can bind LPS directly, TLR4 has the ability to enhance the binding of LPS to MD-2 (Mitsuzawa et al., 2006).

As an example of Gram-positive bacteria, *Mycobacterium tuberculosis* (Mtu) is an inhaled intracellular bacterial pathogen that causes systemic infections in mammals and persists in macrophages of infected organs. Host defense against Mtu is T cell-mediated and requires secretion of IFNγ, TNF, and IL12 and production of reactive nitrogen intermediates (RNIs) as well as signal transduction through MyD88 (Flynn and Chan, 2001; Scanga et al., 2004). The recognition of Mtu is mainly mediated by TLR2 and NOD2 receptor although TLR4 and DC-SIGN also plays a minor role (Kleinnijenhuis et al., 2011).

Furthermore, *Borrelia burgdorferi* (Bbu) is also a model of intracellular microorganism that infects mononuclear cells. It is an agent of Lyme disease, which typically begins with erythema migrans and in addition, might develop cardiac, neurologic and/or arthritic complications (Borchers et al., 2015). Unlike Gram-negative bacteria, Bbu does not contain LPS (Takayama et al., 1987) and instead it has a variety of lipoproteins, many of which embedded on the spirochete's outer membrane (Cervantes et al., 2014). Previous studies have shown that monocytes, DCs, macrophages, NK cells, NK-T cells, and polymorphonuclear cells (PMNs) all contribute to generate a coordinated and robust response to Bbu infection (Moore et al., 2007; Salazar et al., 2003). And the recognition of Bbu signal is mainly modulated by TLR2 and NOD2 receptor (Berende et al., 2010).

In general, the immune response to a bacterial stimulus, particularly sepsis, often comprise the following processes: first, pathogen-associated molecular patterns of the bacterial

pathogens activate pattern recognition receptors of the innate immune system; second, pro-inflammatory and anti-inflammatory mediators are produced in order to contain infection and trigger the clinical signs of sepsis; third, when sepsis and signs of failing organs become obvious, pro-inflammatory processes will stop and a hypo-inflammatory phase with anergy/tolerance of monocytes and apoptosis of T cells will become predominant (Giamarellos-Bourboulis and Raftogiannis, 2012).

Monocytes kill bacteria by producing reactive nitrogen intermediates (RNIs) and reactive oxygen intermediates (ROIs) (Fang, 2004) and through the action of phagolysosomal enzymes (Amer and Swanson, 2002). Although generation of nitric oxide (NO) by human macrophages *in vitro* is difficult to demonstrate, inducible nitric oxide synthase (iNOS) expression in human monocytes can be induced *in vitro* in response to Mtu lipoproteins (Brightbill et al., 1999) and *in vivo* in the lungs of patients with active Mtu (Nicholson et al., 1996), indicating that the RNI-mediated pathway may be operative in human infection. It has been reported that human monocytes and macrophages can kill Mtu in a TLR-dependent and NO-independent way (Thoma-Uszynski et al., 2001).

2.5.2 Fungal pathogens

Candida albicans (Cal), a model of human fungal pathogen, causes mucosal disease as well as systemic infections. A recent study by integrating transcriptional analysis and functional genomics demonstrates that the type I IFN pathway plays a central role in protective immunity mediated by human PBMC against Cal and type I IFNs skewed Cal-induced inflammation from a Th17 response towards a Th1 response (Smeekens et al., 2013).

It is well known that tissue macrophages and neutrophils play crucial roles in defense against fungal infection (Hohl et al., 2006; Romani, 2004). Although recruitment of monocytes to sites of fungal infection has been demonstrated *in vivo* in mouse models (Duong et al., 1998), the underlying mechanism of their contribution to fungal killing is still obscure. Nonetheless, *in vitro* studies of purified murine and human monocytes or cultured macrophages have been performed to characterize the induction of inflammatory and fungicidal mediators, rates of fungal killing (Ibrahim-Granet et al., 2003; Schaffner et al., 1983), and host cell transcriptional responses to *Aspergillus fumigatus* and Cal (Cortez et al., 2006; Kim et al., 2005). But whether *in vitro*-defined mechanisms of fungal inactivation are also functioning *in vivo* and how they contribute to fungal clearance require further investigation.

2.5.3 Pattern-recognition receptors and downstream signaling pathways

The understanding of host defense against pathogenic microorganisms has been revolutionized in the past two decades due to the discovery of dedicated germline-encoded receptors called pattern-recognition receptors (PRRs) that recognize patterns of microbial structures. During infection, the host inflammatory reaction is initiated by the recognition by PRRs of evolutionarily conserved structures of the pathogenic microorganisms known as pathogen-associated molecular patterns (PAMPs) (Netea et al., 2012). So far, five major categories of PRRs have been described: the TLRs, the CLRs (C-type lectin receptors), the NLRs (nucleotide-binding domain-, leucine-rich repeat-containing receptors), the RLRs (RNA helicase RIG-I-like receptors) and the ALRs (cytoplasmic DNA receptor AIM2-like receptors) (Takeuchi and Akira, 2010; Unterholzner et al., 2010).

Among the PRR families, TLRs were the first to be discovered and have been studied most intensively for over 10 years. At the first line of innate host defense, TLRs activate an acute inflammatory reaction after engaging with PAMPs from all the major classes of microorganisms. Subsequently, stimulation via the TLR initiates and modulates the adaptive cellular and humoral immune responses. This requires Toll-IL-1 receptor (TIR) domain-containing adaptors that connect the receptors to downstream effector proteins. Five adaptors have been identified to play a role in TLR signaling: MyD88, MAL (also known as TIRAP), TRIF, TRAM and SARM (O'Neill and Bowie, 2007). Broadly, these adaptors can trigger two main pathways that are dependent on either MyD88 or TRIF, which lead predominantly, but not exclusively, to the production of inflammatory cytokines, or type I IFNs (e.g. IFNα/β), respectively (De Nardo, 2015) (**Figure 2.5.3**).

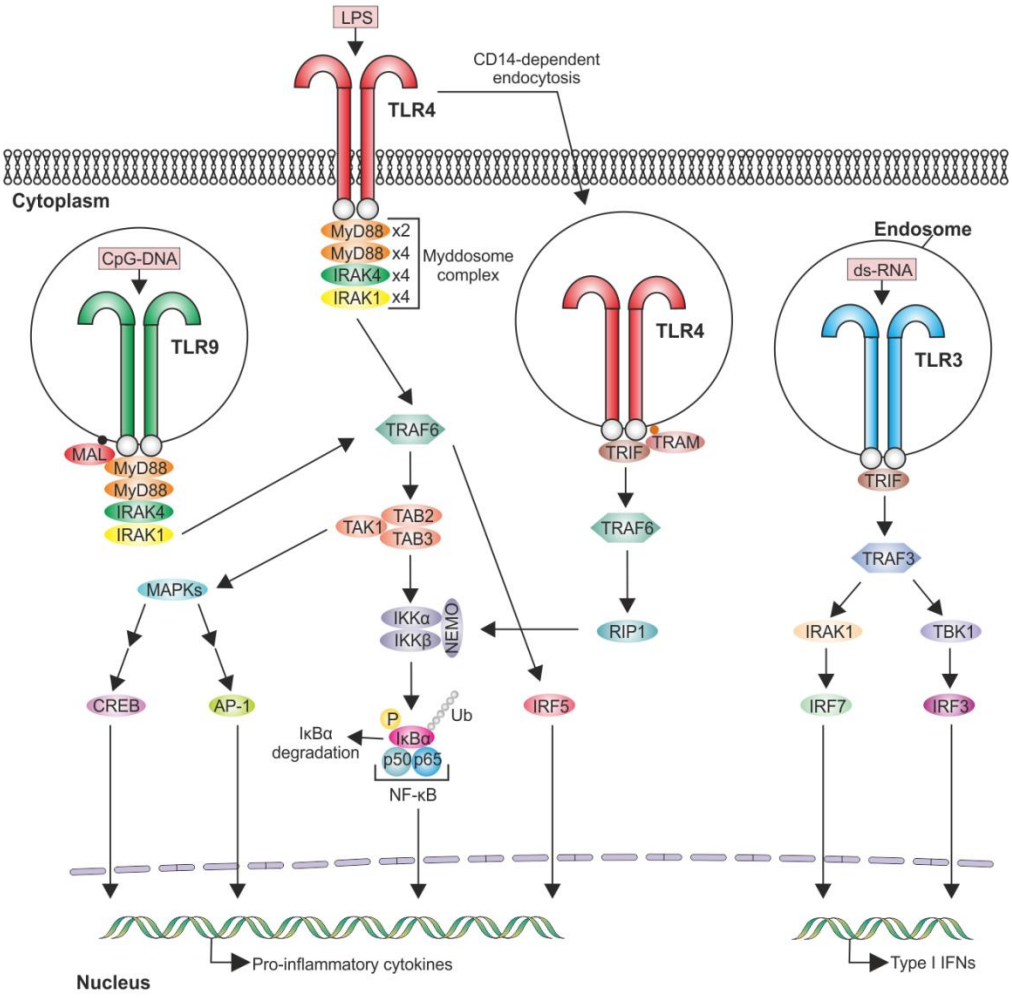

Figure 2.5.3 Overview of TLR signaling pathways. Figure was adapted from (De Nardo, 2015).

MyD88 is the universal signal transducer, because it interacts with almost all the TLRs except TLR3, the receptor for double-stranded RNA. Overexpression of MyD88 (with or without its TIR domain) *in vitro* gives rise to constitutive activation of signaling pathways (Burns et al., 1998). MyD88 directly couples to the intracellular TIR domains of TLR dimers. However, in the case of TLR2 and TLR4, due to suboptimal electrostatic surface charges between their respective TIR domains, MyD88 is thought to require the bridging adaptor, MAL with a more favorably charged TIR domain (Dunne et al., 2003; Fitzgerald et al., 2001; Horng et al., 2002). MyD88 also has a death domain, which recruits IL1 receptor-associated kinases (IRAKs, e.g. IRAK1, IRAK2 and IRAK4), launching signaling pathways that culminate in the activation of transcription factors (TFs), most remarkably NFκB (O'Neill and Bowie, 2007). The multi-protein complex of the MyD88-IRAK family has been called the “Myddosome”. According to

the presence of IRAK1 or IRAK2, two Myddosomes have been characterized. In each, six MyD88 molecules assemble and interact with four IRAK4 molecules, which in turn interact with four IRAK1 molecules or four IRAK2 molecules (Gay et al., 2011; Lin et al., 2010). Furthermore, the Myddosome is also employed by the receptors for IL1, IL18 and IL33, which makes MyD88 especially important for inflammation and host defense. This is also true for the IRAK molecules, with IRAK4 being important for the activation of T cells by IL1 (in TH17 cells) and most likely IL18 (in TH1 cells) (Staschke et al., 2009), which expands the role of this system into adaptive immunity. Activation of the IRAKs allows transient recruitment of the E3 ubiquitin (Ub) ligase, TNFR-associated factor 6 (TRAF6) to the receptor complex. This is followed by the subsequent activation of TRAF6 and its release into the cytosol to form a complex with TAK1, TAB1 and TAB2/3 (Qian et al., 2001b), and thus, activates an IKK complex consisting of NEMO, IKKα and IKKβ, leading to phosphorylation and degradation of IκB, the release and nuclear translocation of NFκB, and the initiation of cytokine production (Napetschnig and Wu, 2013). Additionally, it was observed in mice that TRAF6 signaling also leads to activation of IRF5, which is important for the production of pro-inflammatory cytokines (Takaoka et al., 2005). In parallel, TAK1 triggers the MAP kinase pathway thereby mediating AP-1 and CREB activation, which are also important for cytokine gene transcription (Takeuchi and Akira, 2010). Although the MyD88-dependent pathway predominantly terminates in cytokine production, there is evidence in plasmacytoid dendritic cells (pDCs) that MyD88-IRAK1 signaling could activate the transcription factor IRF7, which initiates the transcription of type I IFNs (Gohda et al., 2004; Honda et al., 2004).

Recruitment of the TRIF triggers a specific signaling cascade that is utilized by both TLR3 and TLR4 (Yamamoto et al., 2003a). Activation of TLR4 induces both the MyD88- and TRIF-dependent pathways. TLR3 can directly interact with TRIF; however, TLR4 requires the bridging adaptor, TRAM, in order to convey signals downstream of TRIF (Yamamoto et al., 2003b). When TLR4 activation initially causes MAL/MyD88 signaling from the plasma membrane, TRAM/TRIF signaling subsequently proceeds due to translocation of TLR4 into endosomal compartments. This was elegantly demonstrated by inhibiting dynamin and clathrin-dependent endocytosis, which had no effect on NFκB but abolished IRF3-dependent IFNβ production (Kagan et al., 2008). TLR4 is degraded within lysosomal compartments when TRIF signaling terminates (McGettrick and O'Neill, 2010). The engagement of TRIF by TLR3 or TLR4 is associated with both TRAF6 and TRAF3 via putative TRAF6 binding motifs to the

receptor complex. The transduction by TRAF3 leads to IRAK1- and TBK-1-mediated activation of IRF7 and IRF3, respectively, and thus results in the production of type I IFNs, whereas TRAF6 recruitment triggers RIP1 kinase and IKK complex, which initiate NFκB activation (Gohda et al., 2004; Sato et al., 2003).

2.6 Whole-genome transcriptional profiling data

Gene expression profiling data are generated by DNA microarray or RNA-sequencing (RNA-Seq) technology, which simultaneously measures the relative activity of thousands of previously identified target genes, to provide a global view of the complete transcriptome for cell of interest.

The core principle for microarray technology is hybridization between two DNA strands, i.e. the property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs. Each chip consists of thousands of DNA spots. Each DNA spot contains picomoles (10-12 moles) of a specific DNA sequence (probe), which can be a short fragment of a gene or other DNA element that are used to hybridize a complementary DNA (cDNA) or cRNA (target) under high-stringency conditions. After washing off non-specific bonding sequences, only strongly paired strands with high numbers of complementary base pairs will remain hybridized. Signals can be detected from fluorescently labeled target sequences that bind to a probe sequence. The signal intensity from a spot relies on the amount of target sample binding to the probes present on that spot and can be quantified by image analysis (Staal et al., 2003). One example for Affymetrix microarrays is illustrated in **Figure 2.6**.

Compared to one-color microarrays, two-color microarrays can be performed by using cDNA from two samples (e.g. cancer cells versus normal cells) labeled with two different fluorophores. The commonly used fluorescent dyes include Cy3 (a fluorescence emission wavelength of 570 nm, green) and Cy5 (a fluorescence emission wavelength of 670 nm, red). The two Cy-labeled cDNA samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores after excitation with a laser beam of a defined wavelength. Relative intensities of each fluorophore may then be used in ratio-based analysis to identify up-regulated and down-regulated genes (Tang et al., 2007).

Since the detection of expression levels in microarrays is dependent on known gene sequences or existing reference genomes or transcriptomes, the disadvantage of this technology is that it is unable to identify novel gene targets or other molecular fragments such as unknown microRNAs. In recent years, the development of RNA-Seq technologies enables a fast way to characterize and quantify gene expression from a whole transcriptome. Independent of a reference genome or transcriptome, RNA-Seq can also be used for new model organisms whose genome has not been sequenced yet, which microarray is unable to accomplish. Until recently, researchers still choose the microarray technology preferentially due to lower costs especially with large sample size. Eventually however, RNA-Seq technologies will subsidize microarray technologies completely.

During the last 10 years, the two dominant microarray manufacturers in the world were Affymetrix (Santa Clara, California, USA) and Illumina (San Diego, California, USA). Affymetrix uses a silicon chip as solid surface that the probes can be fixed whereas Illumina uses Nano beads instead of the large solid support. In 2015, Illumina has stopped its microarray product for assessing gene expression in mice and suggested to its customers to use RNA-Seq instead.

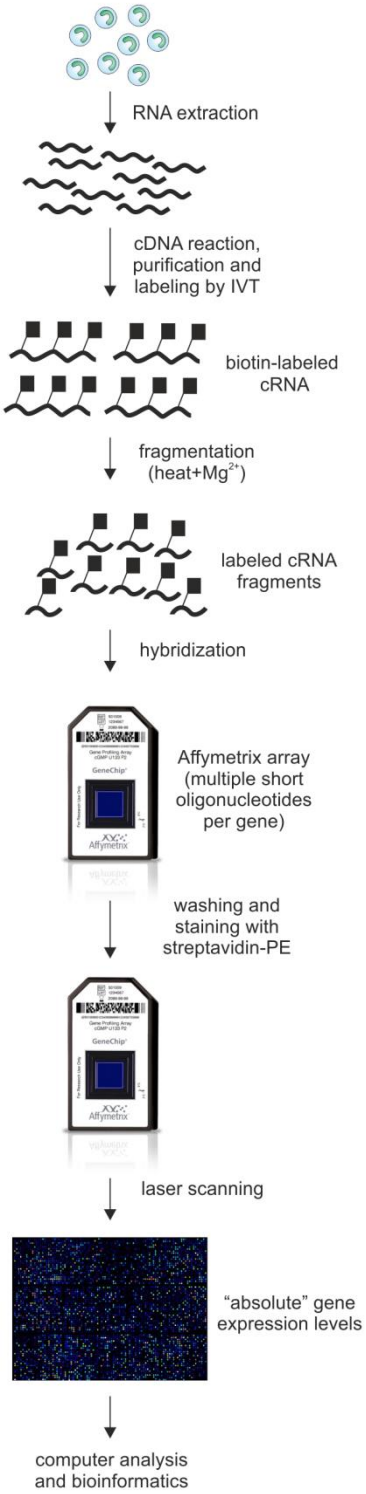

Figure 2.6 The principle steps of Affymetrix microarray experiments. Figure was adapted from (Staal et al., 2003). Total RNA is extracted from one cell population and cDNA is prepared for each array chip. The cDNA is used in an *in vitro* transcription (IVT) reaction to generate biotin-labeled cRNA. After fragmentation with heat and Mg^{2+} ions, this cRNA is hybridized to microarrays, washed and stained with PE-conjugated streptavidin. And subsequently the stained chip is scanned on a laser scanner. The signal intensity of each probe on the microarray represents the "absolute" gene expression level which is finally analyzed by bioinformatics methods.

2.7 Public transcriptomic datasets of immune cells

Transcriptomics has considerably contributed to a better understanding of immune cell function and regulation. Large consortia such as the ImmGen consortium (Best et al., 2013; Bezman et al., 2012; Cohen et al., 2013; Gautier et al., 2012; Miller et al., 2012) or the Human Immunology Project Consortium (Poland et al., 2013) compiled extensive datasets and defined a core transcriptional program for murine tissue macrophages and dendritic cells (DC) under steady state conditions (Gautier et al., 2012; Miller et al., 2012). A complementary approach has been introduced by InnateDB (Breuer et al., 2013). Data on molecular interactions between proteins of the innate immune system derived from smaller datasets have been compiled and can be used to reveal mechanistic insights into immune cell function (Hume et al., 2010; Mabbott et al., 2010). Applying new sequencing technologies (a combination of RNA-Seq, ChIP-Seq and ATAC-Seq data analysis) enables to study epigenetic landscapes of murine tissue-resident macrophages (Lavin et al., 2014).

Unfortunately, meta-analysis of small datasets has been hampered by several challenges. Differences in the genetic background of mice, in stimulation conditions, and the combination of *in vitro* and *in vivo* data limit or even bias model generation of incongruous data sets (Mabbott et al., 2010). Moreover, comparative studies have identified substantial differences in immune cell gene expression between mice and humans (Schroder et al., 2012; Shay et al., 2013). Therefore, it remains to be fully elucidated, how immune cell activation - particularly in human macrophages - is transcriptionally controlled and to which degree these pathways are conserved across species (Murray and Wynn, 2011). Standardizing data acquisition and assembling larger datasets, such as by the ImmGen consortium (Heng and Painter, 2008), is necessary to answer such questions.

2.8 Computational approaches applied to gene expression profiling

In the next paragraphs, I will introduce mainly those computational approaches that were utilized in the four major projects described in this thesis. The field of gene expression profiling has seen an enormous development during the last decade and several approaches have become standard, while others are still experimental. Efforts to standardize and compare methodologies from the DREAM (Dialogue for Reverse Engineering Assessments and Methods) competitions and the tools compared in DREAM5 have also been used as part of this thesis.

2.9 Dimensionality reduction

As high-throughput data consist of thousands of variables (genes), that means, thousands of dimensions, which cannot be recognized by humans. Hence, reduction of dimensionality is necessary to visualize the fundamental structures within a dataset. There are many ways to reduce dimensionality of data. The following two, principal component analysis and correlation analysis, are two widely used computational methods.

2.9.1 Principal component analysis

Principal component analysis (PCA) is a linear transformation method that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is always less than or equal to the number of original variables, which consequently reduces the dimensionality of the original data. The transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. PCA is sensitive to the relative scaling of the original variables (Jolliffe, 2002).

As PCA is a linear transformation that focuses on keeping the low-dimensional representations of dissimilar data points apart from each other, it has limitations to keep the low-dimensional representations of similar data points close together. Therefore, non-linear dimensionality reduction techniques have been developed, for example, t-SNE, to visualize the similarities within a high-dimensional dataset (van der Maaten and Hinton, 2012).

2.9.2 Correlation analysis

In statistics, correlation refers to statistical relationships between two dependent variables. There are several correlation coefficients, often denoted ρ or r, measuring the degree of correlation. A perfect correlation of 1 or -1 occurs when each of the variables is a perfect monotone function of the other. Similarly, when the correlation coefficient equals to 0, the corresponding two variables are completely independent of each other.

2.9.2.1 Pearson correlation coefficient

The most widely used correlation coefficient is the Pearson correlation coefficient, which is a measure of linear correlation between two variables. The values are always between 1 and -1 inclusive, where 1 is the total positive correlation, 0 is no correlation, and -1 is the total negative correlation. It was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s (Myers et al., 2010).

Pearson's correlation coefficient between two variables *X* and *Y* is defined as the covariance of the two variables divided by the product of their standard deviations.

$$\rho = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(\mathrm{y_i} - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

where:

- x_i and y_i are values from *X* and *Y*, respectively; n is the size of the samples
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (sample mean)

2.9.2.2 Spearman's rank correlation coefficient

Other correlation coefficients such as the Spearman's rank correlation coefficient have been developed to be more robust than the Pearson correlation – that is, more sensitive to nonlinear relationships. Spearman's coefficient is appropriate for both continuous and discrete variables, including ordinal variables.

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. For a sample of size n, the n raw scores $\mathrm{X_i}$, $\mathrm{Y_i}$ are converted to ranks $\mathrm{x_i}$, $\mathrm{y_i}$, and ρ is computed from:

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(x_i - \mathrm{y_i})^2}{n(n^2-1)}$$

Identical values (rank ties or value duplicates) are assigned a rank equal to the average of their positions in the ascending order of the values (Myers et al., 2010).

2.10 Cluster analysis

Cluster analysis is unsupervised machine learning to group and dissect data objects based only on information retrieved in the data that describes the objects and their relationships. The objects within a group should be similar (or related) to one another and different from (or unrelated to) the objects in other groups after clustering. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better or more distinct the clustering will be. Either for understanding or as a utility, cluster analysis has long played an important role in a wide variety of fields such as psychology, biology, statistics, pattern recognition, information retrieval, machine learning, and data mining. Following are some examples of clustering methods that are commonly applied in bioinformatics.

2.10.1 Hierarchical clustering

Hierarchical clustering is an unsupervised learning approach which aims to construct a hierarchy of clusters. In general, there are two types of strategies or algorithms for hierarchical clustering. One is agglomerative (a "bottom up" approach): each observation starts in its own cluster, and pairs of closest clusters are merged as one moves up the hierarchy. The other is divisive (a "top down" approach): all observations start in one cluster, and separation of discriminate observation is performed recursively as one moves down the hierarchy (Maimon and Rokach, 2010).

The results of hierarchical clustering are usually displayed as a dendrogram. An appropriate metric as a measurement of dissimilarity between sets of observations is required to decide which clusters should be combined, or where a cluster should be split. The commonly used metrics for hierarchical clustering are Euclidean distance, Squared Euclidian distance, Manhattan distance, maximum distance and Mahalanobis distance (Maimon and Rokach, 2010).

2.10.2 K-means clustering

K-means clustering is a method of vector quantization, a machine learning approach that was originally proposed by James MacQueen in 1967 through the idea of Hugo Steinhaus in 1957 (Manning et al., 2008). The aim of this method is to partition n observations into a certain number (k) of clusters in which each observation belongs to the cluster with the

nearest mean, serving as a prototype of the cluster. The algorithm can be described as following:

Given a set of observations $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k $(\mathrm{k} \leq \mathrm{n})$ sets $S = (\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_\mathrm{n})$, so as to minimize the within-cluster sum of squares. In other words, its objective is to find:

$$\arg\min \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{x} \in \mathrm{S}_\mathrm{i}} ||\mathrm{x} - \mu_\mathrm{i}||^2$$

where μ_i is the mean of points in S_i.

The problem is computationally difficult, i.e. it is a non-deterministic polynomial-time hard (NP-hard) problem. However, if k and d (the dimension) are fixed, the time complexity of the problem would be $\mathrm{O}\left(\mathrm{n}^{\mathrm{dk}+1}\mathrm{logn}\right)$ (Wu, 2012).

2.10.3 Self-Organizing Map

Self-Organizing Map (SOM) or Kohonen Self Organizing Feature Map was a vector quantization method invented by Teuvo Kohonen (Kohonen, 2013). It is an efficient data compression technique for visualizing low-dimensional views of high-dimensional data. It has been adopted as a standard analytical tool in many fields of science such as statistics, signal processing and systems biology.

The stages of the SOM algorithm can be summarized as follows. First, initialization: choose random values for the initial weight vectors w_j. Second, sampling: draw a sample training input vector x from the input space. Third, matching: find the winning neuron $I(x)$ with weight vector closest to input vector. Fourth, updating: apply the weight update equation $\Delta w_{ji} = \eta(t) T_{j,I(x)}(t)(x_i - w_{ji})$. Fifth, continuation: keep returning to step 2 until the feature map stops changing (Kohonen, 2013).

However, studies in the last two decades indicate that SOM has limitations for either clustering/vector quantization (VQ) or multi-dimensional scaling (MDS) when comparing to traditional VQ or MDS techniques (Balakrishnan et al., 1994; Bezdek and Nikhil, 1995). SOM's ability of doing both VQ and MDS at the same time is also challenged by a new combined

approach, online K-means clustering plus Sammon mapping of the cluster centroid (Flexer, 1997).

2.10.4 Markov Cluster (MCL) Algorithm

The MCL algorithm is a fast and scalable unsupervised cluster algorithm for networks based on simulation of stochastic flow in networks (Van Dongen, 2008). The algorithm was invented by Stijn van Dongen and fully explained in his PhD thesis.

The engine of the MCL algorithm is a Markov chain process, which is a discrete uncoupling process for finite space. The algorithm consists of three main steps: parameter setup, MCL process computing and result interpretation. The MCL process takes a stochastic matrix as input, and then alternates expansion and inflation, each step defining a stochastic matrix in terms of the previous one. Given a graph (or a network) G, the algorithm employs the process by applying it to the matrix of random walks on G (Van Dongen, 2008).

The MCL algorithm has following properties: First, it generates well-balanced flat (non-hierarchical) clusters. Second, it is intrinsically a bootstrapping method, by which seeding information, especially the number of clusters, is not required. Third, it has a natural parameter (inflation) affecting cluster granularity (size). Fourth, it is scalable to sparse graph/matrix implementation techniques. Fifth, mathematical results link MCL process iterands, the cluster interpretation, inflation, and the number of clusters together (Van Dongen, 2008).

2.11 Network analysis on high-throughput data

Since 15 years, the availability of high-throughput gene expression data has made it possible to infer large-scale gene regulatory networks. The gene-to-gene relationships could be computed by applying a variety of mathematical metrics given by a plenty of sample size (in some cases even larger than 100 samples) to include individual variations for statistical predictions (Margolin et al., 2006). Due to this prerequisite, currently gene expression profiles from microarray experiments are the major resource for gene regulatory network inference rather than using high-resolution sequencing data to control the research costs. The inferred gene-to-gene relationships or gene pairs are often ranked by certain score defined by each algorithm and the high-ranking gene pairs are considered as gene-gene interactions, which can be visualized by network analysis tools in 2D or 3D fashion.

Cytoscape (http://www.cytoscape.org/) is an open source software written in Java for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other high-throughput data. Although it was originally designed for biological research and made public in 2002, now it is a general platform for complex network analysis and visualization. Cytoscape core distribution provides a basic set of features for data analysis, integration and visualization and also additional plugins (Apps) for customized usage such as network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases. Most of the Apps are freely available from Cytoscape App Store (http://apps.cytoscape.org/).

There have been various attempts to reconstruct gene regulatory networks from microarray expression data. Of them, the commonly applied algorithms will be introduced.

2.11.1 Network inference algorithms

A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Bayesian network is one of the first implemented algorithms to infer gene regulatory networks from high-throughput microarray expression data 15 years ago (Friedman et al., 2000). However, network prediction has always been challenged due to the limited amount of independent experimental conditions and noise inherent in the measurements. Therefore, supervised approaches were used to systematically integrate expression data and biological prior knowledge, related, for instance, to transcription factor binding locations in promoter regions or partially known signaling pathways from the literature (Werhli and Husmeier, 2007).

As the most popular manner to represent statistical relationships between two observations, correlation methods have been applied to analyze gene expression data in the field of systems biology at the very beginning. BioLayout Express ^{3D} is one of the fast and convenient network generation tools on the basis of correlation methods (Theocharidis et al., 2009). However, it has been demonstrated that the gene expression relationship between a TF and

its targets is complex. In most cases, they do not have a correlated expression profile over a time course. Sometimes, in fact, there is a lag time between the expression of the TF and its target (Qian et al., 2001a). Moreover, correlation methods have shown bias in predicting different network motifs, e.g., they are good at inferring feed-forward loops, but not knockouts and directionality (Marbach et al., 2012).

Mutual information, another metrics for measurement of dependencies between variables, was utilized to calculate the relationships between each gene pair within gene expression profiles since 2000 (Butte and Kohane, 2000). Unlike correlation, mutual information does not assume linearity, continuity, or other specific properties of the dependence. As such, mutual information possesses the flexibility to detect regulatory interactions that might be missed by linear measures such as the correlation coefficient.

Formally, the mutual information of two discrete random variables X and Y can be defined as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

Or in the case of continuous random variables, the summation is replaced by definite double integral (Mc Mahon et al., 2014):

$$I(X;Y) = \int_{Y} \int_{X} p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right) dxdy$$

Due to these advantages, several network prediction methods based on MI have been released during the last decades. The more advanced method ARACNe (algorithm for the reconstruction of accurate cellular networks), combining mutual information and a well-known staple of data transmission theory, the 'data processing inequality' (DPI), was first reported to be successfully applied on gene-expression profile data from human B cells (Basso et al., 2005; Margolin et al., 2006). ARACNe has been also extensively applied to time-course data to develop TimeDelay-ARACNE and was tested on *Saccharomyces cerevisiae* and *E. coli* networks (Zoppoli et al., 2010). Furthermore, the context likelihood of relatedness (CLR) algorithm also uses mutual information for scoring the similarity between the expression levels of two genes in a set of microarrays and applies an adaptive background correction step to eliminate false correlations and indirect influences (Faith et al., 2007).

These methods based on information theory were proposed to be better performed in predicting gene regulatory networks of large complex mammalian cells (Basso et al., 2005).

Regression techniques, which fit the data to *a priori* models, are one big category of network inference methods. For example, Yeung *et al.* uses singular value decomposition to construct a family of candidate solutions and then uses robust regression to identify the solution with the smallest number of connections as the most likely solution (Yeung et al., 2002). More recently, GENIE3 (GEne Network Inference with Ensemble of trees) was proposed as the best performer in DREAM 4 Project. It decomposes the prediction of a regulatory network between p genes into p different regression problems, in each of which the expression pattern of one of the genes (target gene) is predicted from the expression patterns of all the other genes (input genes), using tree-based ensemble methods random forests (Huynh-Thu et al., 2010).

2.11.2 Community-based methods

The results of network inference rely predominantly on the studied data including type of the data (real or simulated), size of the network, number of samples, amount of noise, experimental design (observational, experimental or interventional), type of network structure (random, scale-free or small-world), etc. Due to these factors, there is no one "right" method that is able to fulfill all different biological, technical and experimental conditions best (Emmert-Streib et al., 2014), but it has been shown that a number of methods have the potential to infer similar biological information (de Matos Simoes et al., 2013). Therefore, a recent trend in this field is to use ensemble methods (Zhang and Singer, 2010) or community-based methods to improve the stability and accuracy of the inferred networks. Computational studies from the DREAM project demonstrate that community-based methods integrated from many individual approaches are much more robust than individual methods (Marbach et al., 2012). In brief, the underlying idea is to first apply individual inference methods and then aggregate all separate outcomes by average rank to build a consensus network. So far, these methods have been applied to gene expression data derived from microorganisms only and the results were evaluated based on known biological interactions (Marbach et al., 2012).

2.11.3 Network analysis on immune cells

Recently, several elegant studies have revealed the importance of analyzing networks based on gene expression profiling in immune cells such as macrophages (Martinez et al., 2006; Nau et al., 2002; Ramsey et al., 2008) and Th17 cells (Ciofani et al., 2012; Yosef et al., 2013). By combining unsupervised learning of transcriptome data and supervised learning of database-stored information, these studies show how technological and analytical advances can reveal network structures in immune cells. But most mathematical approaches require large datasets. For example, the reverse engineering algorithm ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks), which is robust in predicting gene regulatory networks for mammalian cells, has previously been employed to characterize B cell activation (Basso et al., 2005) and further refined during the last few years (Marbach et al., 2012). The requisite of these algorithms is to have at least 100 gene expression profiles as input. However, due to the lack of large enough datasets, reverse engineering approaches were not applied to other immune cells like human macrophages and monocytes so far prior to this thesis.

2.12 Knowledge-based analysis

So far, only data-driven approaches have been introduced. However, it is also very important to link the data to prior knowledge from literature or databases. Therefore, the most commonly used knowledge-based analytic methods will be described in the next paragraphs.

2.12.1 Gene Ontology (GO) and GO enrichment analysis

The Gene Ontology (GO) project (http://geneontology.org/), founded in 1998, is a collaborative effort that provides controlled vocabularies of defined terms representing gene product properties. The aim of the project is to address consistent descriptions of gene products across databases. The ontology covers three domains: first, cellular component, the parts of a cell or its extracellular environment where a gene product is involved; second, molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis; and third, biological process, operations or sets of molecular events with a defined start and end, relevant to the functioning of integrated living units: cells, tissues, organs, and organisms.

The ontology is structured as a directed acyclic graph (hierarchical structure) where each term has defined relationships to one or more other terms in the same domain, and sometimes to other domains. The GO vocabulary is designed to be species-independent, and includes terms applicable to prokaryotes and eukaryotes, and single and multi-cellular organisms.

A GO enrichment analysis will find which GO terms are over-represented or under-represented using annotations for a given set of candidate genes, e.g. genes that are up- or down-regulated under certain conditions. The over-representation of GO terms could be evaluated by several values, such as enrichment p-value. P-value is the probability of observing at least x number of genes out of the total n genes in the list annotated to a particular GO term, given the proportion of genes in the whole genome that are annotated to that GO Term, i.e. the GO terms shared by the given genes are compared to the background distribution of annotation. The closer the p-value is to zero, the more significant the particular GO term is associated with the set of genes.

There are a number of different GO enrichment tools. Tools may use diverse algorithms and perform different statistical tests. In this study, I mainly used a Cytoscape plug-in BiNGO (see more in **Materials and Methods**).

2.12.2 KEGG Pathway and pathway enrichment analysis

KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.genome.jp/kegg/) is a database resource for understanding high-level functions of the biological system, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

Similar to GO enrichment analysis, pathway enrichment analysis aims to identify over-represented KEGG pathways using a given set of candidate genes.

3 Materials and Methods

3.1 Gene expression profiling data generation and retrieval

3.1.1 Illumina microarrays

For human macrophage resource data (GEO accession number: GSE47189), total RNA was further purified using the MinElute Reaction Cleanup Kit (Qiagen). Biotin labeled cRNA was generated using the TargetAmp Nano-g Biotin-aRNA Labeling Kit for the Illumina System (Epicentre). Biotin labeled cRNA (1.5 µg) was hybridized to HumanHT-12V3 or HumanWG-6V3 Beadchips, and scanned on an Illumina HiScanSQ system. Raw intensity data of a total number of 384 samples were processed in Genome Studio (Illumina) excluding probesets with missing bead types to increase validity. The microarray experiments were performed mainly by Susanne V. Schmidt, Michael R. Mallmann and Marc Beyer.

The human PBMC microarray data (GEO accession number: GSE42606) were retrieved from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). The dataset comprises 299 samples generated by Illumina HumanHT-12 V4.0 expression Beadchips.

3.1.2 Affymetrix microarrays

The murine yolk sac and tissue macrophage data were generated by Affymetrix MoGene-2_1-st-v1 arrays. In this research, there are in total 27 cell populations, each of which has at least two or three biological replicates (sample size = 81). The microarray experiments were performed by a collaboration partner from KFB Regensburg (Kompetenzzentrum für Fluoreszente Bioanalytik, Regensburg).

3.2 Primary expression data handling

3.2.1 Illumina microarrays

For each study, the complete transcriptomic dataset was imported into Partek Genomics Suite (PGS) for further analysis including quantile normalization. Batch effects coming from different array experiments were removed from normalized log2-transformed data. Background signal was calculated within R based on coefficient of variation. The corresponding log2-transformed background value will be described for each dataset in the results section. Genes are only kept for further analysis if their mean expression values are higher than background in at least one group or condition. Afterwards, multi-probes were

filtered allowing only one probe with the highest mean expression within each group representing corresponding genes. Only unique present genes, which represent most informative genes, were retained for further analyses such as network analysis.

3.2.2 Affymetrix microarrays

Robust Multi-array Average (RMA) function was used to normalize the data during importation of raw Affymetrix data into PGS. Background value was set to be the median of the entire normalized data throughout genes. Probes below the background and multiple probes were filtered as described for the Illumina data (see **Chapter 3.2.1**). For the murine tissue macrophage data, 17,888 unique present genes were derived for further analysis.

3.3 Identification of differentially expressed genes and hierarchical clustering

Differentially expressed (DE) genes were defined by Analysis of Variance (ANOVA) models by setting certain fold change (FC) and False Discovery Rate (FDR, Benjamini and Hochberg, 1995) adjusted p-value cutoffs to determine differences between two comparable groups or conditions. The corresponding cutoffs used for a particular comparison will be described in the result section. The most variable genes were determined in ANOVA models to compare the variations throughout groups by t-test, i.e. genes with most significant p-values. Hierarchical clustering using Euclidean distance on both genes and samples enables visualization of DE genes or most variable genes on particular conditions.

3.4 Co-regulation analysis by BioLayout Express^{3D}

BioLayout Express^{3D} (BioLayout) is a powerful tool for the visualization and analysis of network graphs based on co-expression (Theocharidis et al., 2009). In this PhD work, BioLayout was applied to either discriminate different sample groups or identify gene-gene co-expression interactions. Sample-sample correlation or gene-gene correlation was calculated and transformed into 3D coordinates. A specific Pearson correlation cutoff is required to construct and visualize the co-regulation network. For gene-wise co-regulation networks, a proper Pearson correlation cutoff should be chosen to fulfill the power law for degree-node distribution.

3.5 Self-Organizing-Map (SOM) clustering

To reduce the dimensionality of high-dimensional data for visualization, SOM clustering was used. It projects the input space on prototypes of a low-dimensional regular grid that can be effectively utilized to explore properties of the data (Kohonen, 1982). SOM clustering was performed to classify the different conditions using PGS. First, the expression values were z-transformed (z-scores) and this was followed by 20,000 training iterations to cluster similar probes close to each other on the map. In our settings, the whole transcriptome was divided into 10×10 clusters, and the expression values of each cluster gene are rescaled to one eigenvalue, which represent the general expression value of this cluster. The resulting data are then visualized as a heatmap representing increased values in red, decreased values in blue and intermediate values in green. It needs to be mentioned that the input data (e.g. number of samples or conditions) influences the cluster structure and standardized mean expression values (z-scores).

3.6 Pearson and Spearman correlation coefficient matrices

To investigate the similarities across sample groups, Pearson correlation coefficients or Spearman correlation coefficients in a pairwise fashion between all conditions were computed using PGS, which results in a Pearson or Spearman correlation coefficient matrix. The individual samples were grouped into one using mean expression values of the corresponding group. The resulting correlation coefficient matrices were visualized as heatmaps with or without hierarchical clustering of the sample groups.

3.7 Calculation of the vectors for the multi-dimensional model of macrophage activation

The multi-dimensional model of macrophage activation was established by grouping the samples according to the clusters obtained by the PCCM analysis, utilizing the 3D coordinates of the individual macrophage samples determined by co-regulation analysis, calculating mean vectors for the clusters and plotting the information in a 3D graph using the coordinates of the baseline macrophages (M^b) as the origin. The coordinates of the nodes can be joined by conditions or clusters using 'Collapse Nodes by Class' function in BioLayout by setting M^b as origin (0, 0, 0) and then the joined coordinates of other conditions or

clusters are rescaled based on the origin in 3D space. The vectors starting from M^b to all activation states were plotted in 3D space using Matlab.

3.8 Gene modular analysis by weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) can be used to identify underlying data structures in a complex dataset (Langfelder and Horvath, 2008). WGCNA was used to identify co-regulated genes associated with the studied conditions. The WGCNA R package (http://labs.genetics.ucla.edu/horvath/htdocs/CoexpressionNetwork/ Rpackages/WGCNA/) was utilized for the analysis. The mean expression matrix of 29 conditions from preprocessed human macrophage activation data was used as input. The standard parameters such as power and minimal module size were altered to obtain better clustering of gene modules. For each module the eigengene corresponding to the first principal component of a given module was calculated. The network for each module of interest was generated using the "1-TOMsimilarityFromExpr" function of the WGCNA R package. The eigengenes of 49 identified modules was correlated to the 29 conditions and the correlation matrix was visualized as a heatmap. The WGCNA analysis was performed mainly by Thomas Ulas.

3.9 GO enrichment analysis and GO network visualization

To link the data to prior knowledge, GO enrichment analysis was applied by using the Cytoscape plug-in BiNGO (v2.44) (Maere et al., 2005). To include only significant results, the FDR q-value threshold was set to 0.05 as default. The Cytoscape plugins Enrichment Map (v1.1) (Merico et al., 2010) and Word Cloud (Oesper et al., 2011) were used to visualize the GO networks. Certain cutoffs for similarity of the GP terms, Jaccard coefficient (default: 0.25), and their significance FDR q-value (default: 0.1) were utilized depending on the size of the networks. It will be described in the result section if the default settings were not used.

3.10 Pathway enrichment analysis

To understand the involvement of a certain list of candidate genes in biological pathway, pathway enrichment analysis was performed in PGS, which derives well-established pathway information from KEGG database (http://www.genome.jp/kegg/) and compute the enrichment scores and enrichment p-values (t-test) for each particular pathway based on corresponding gene lists.

3.11 miRNA-Seq data generation and analysis

Next-generation sequencing of miRNAs for four human macrophage conditions, i.e. baseline (M^b), IFNγ ($M^{IFNγ}$), IL4 (M^{IL4}) and the combination of TNF, PGE_2 and P3C (TPP, M^{TPP}), was performed according to the manufacturer's recommendations. In brief, $5x10^6$-$2x10^7$ macrophages were harvested and total RNA including small RNAs was isolated. Small RNA libraries were generated from 1 µg total RNA with the TruSeq Small RNA Sample Preparation Kit (Illumina). After successful ligation of 3' and 5' adapters to RNA molecules, RNA was reverse-transcribed using SuperScript II reverse transcriptase (Invitrogen). cDNA was amplified by 11 PCR cycles with high-fidelity Phusion Polymerase (Finnzymes). cDNA with the size of miRNAs plus ligated adapters was purified on a pre-cast 6% Tris/Borate/EDTA polyacrylamide gel electrophoresis gel (Invitrogen). Generation of clonal clusters from single molecules of the cDNA library was done with the TruSeq Cluster Kit (Illumina) on a CBot station. Sequencing by synthesis was performed by using the TruSeq SBS Kit on a HiScanSQ system (Illumina). The sequencing experiments were performed mainly by Susanne V. Schmidt. Afterwards, sequencing reads were retrieved as FASTQ files. After demultiplexing adapter sequences were trimmed from each read using Flicker 3.0 (Illumina). Trimmed reads were mapped to the human genome hg19 and hairpin and mature human miRNAs deposited in miRBase version 19 using the short read aligner Bowtie 0.12.9 (Langmead et al., 2009) with no mismatch allowance. The number of reads mapping to a specific miRNA sequence were counted within PGS. The dataset was then normalized by using the statistical software R package DESeq (Anders and Huber, 2010) and miRNAs having less than one normalized read count in all M^b samples were excluded. The read counts were transformed into log2 counts per million (cpm) and were divided by the corresponding library size (in millions) by using the R package limma (Smyth, 2005). The R package sva (Johnson et al., 2007) was used to perform a batch removal for the random factors date and donor. Then miRNAs having less than one transformed read count in all samples of the same condition were excluded. Differentially expressed miRNAs between $M^{IFNγ}$, M^{IL4} or M^{TPP} were determined against M^b by using the R package limma (Smyth, 2005) with a p-value of 0.05 as well as an absolute fold change of 2 as cutoffs. Finally, for each condition a set of uniquely differentially expressed miRNAs was determined, which was then sorted by the transformed expression values. The first five most highly abundant up-regulated and the first five most highly abundant down-

regulated miRNAs were chosen to be represented within a heatmap. The analysis of miRNA-Seq data was performed mainly by Martina Emde and Jil Sander.

3.12 Gene set enrichment analysis (GSEA)

GSEA was applied to discover similar gene expression signatures between a known list of condition-specific genes and two groups of interest, especially between two datasets. From one unexplored dataset, every time two groups, i.e. experimental condition versus control, were taken to be compared by only looking into a given gene set. For example, this approaches was used to discover whether the transcriptional programs identified in *in vitro* activated macrophages can be also found in *in vivo* tissue macrophages. Two microarray datasets generated from human alveolar macrophages were compiled containing three clinical groups, smokers, non-smokers and chronic obstructive pulmonary disease (COPD) patients. Smokers and COPD patients were compared with non-smokers separately to be able to apply the GSEA algorithm. The 49 WGCNA gene modules (see **Chapter 3.8**) defined from the human macrophage dataset representing stimulus-specific gene signatures were employed as 49 gene sets. 1,000 permutations were performed for each gene set in PGS (Subramanian et al., 2005). In the end, normalized enrichment score (NES), allowing comparisons of overrepresentation between different gene sets, together with enrichment p-values of GSEA were visualized as Volcano plots.

3.13 Gene regulatory network predictions

The filtered expression matrix (with unique present genes) from the macrophage activation dataset (Xue et al., 2014) were utilized as input for network inference methods. The following are the five major network inference methods that were used.

3.13.1 Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe)

ARACNe is a robust reverse engineering algorithm to reconstruct mammalian cellular networks from realistic synthetic data and microarray data (Margolin et al., 2006). The expression data were loaded into an integrated genomic analysis platform geWorkbench version 2.5.1 (http://wiki.c2b2.columbia.edu/workbench/index.php/Home) to implement the ARACNe algorithm for network analysis. The algorithm includes two major steps: one is to identify gene pairs exhibiting correlated transcriptional responses by measuring mutual information (MI) or all input pairs of genes i and j within gene expression profiles (see also

Chapter 2.11.1); the other is to eliminate indirect interactions by applying data processing inequality (DPI) (Basso et al., 2005; Margolin et al., 2006). All genes present in all array samples were taken into calculation with DPI tolerance 0.1 and the significance threshold for p-value was 1e-7. Gene pairs were ranked by their MI scores.

3.13.2 TINGe

Since the introduction of ARACNe several improvements of the original algorithm and novel algorithms have been introduced for reverse network engineering of transcriptome datasets. To ensure robustness of the computed network, a second reverse engineering algorithm, namely the TINGe (Tool for Inferring Network of Genes) algorithm (Aluru et al., 2013) was applied. Written in C++ TINGe is also based on information theory. To compare results from ARACNe and TINGe derived networks, the same significance threshold ($p < 1e-7$) and the same DPI tolerance (0.1) were used.

3.13.3 Context Likelihood of Relatedness (CLR)

CLR is one of the most commonly used approaches for gene regulatory network inference that have been integrated into the GP-DREAM web tool (Marbach et al., 2012; Reich et al., 2006). A gene x condition matrix of expression values can be uploaded into GP-DREAM. The CLR algorithm progresses through two main steps. First, a matrix of MI values is calculated for all input gene pairs using B-spline smoothing (spline degree of 3) and the gene expression values were discretized into 10 bins. Second, CLR estimates the significance of a gene pair by comparing the corresponding MI values to an empirically defined background distribution of MI values. The significance of a gene pair is defined by a modified z-score (Faith et al., 2007). Gene-gene connection predictions are ranked according to the modified z-score. A cutoff of 100,000 edges was used to obtain high ranking inferred interactions.

3.13.4 Gene Network Inference with Ensemble of Trees (GENIE3)

GENIE3 decomposes the prediction of a regulatory network between n genes into n different regression problems. In each of the regression problems, the expression pattern of one of the target genes is predicted from the expression patterns of all other genes based on Random Forests (Huynh-Thu et al., 2010). The importance of one gene in the prediction of the target gene is taken as an indication of a putative regulatory link. 100,000 putative regulatory links are then aggregated over all genes to provide a ranking of interactions to reconstruct a network.

3.13.5 BasicCorrelation

Pearson's correlation was employed to compute the relationships between all gene pairs within gene expression data. This method is also compiled within GP-DREAM. The relationship between gene x and y were ranked via Pearson's correlation coefficient (Marbach et al., 2012). Again, only 100,000 gene pairs with highest correlation were kept.

3.13.6 Consensus network

All gene-gene interactions derived from each individual network prediction were ranked by its ranking function. By computing average rank for each gene pair, 10,000 top ranking interactions was obtained. These putative gene-gene interactions were imported into Cytoscape (Saito et al., 2012) to build a consensus network. Duplicated edges and self-loops were removed.

3.13.7 Network topology analysis

Networks were visualized in a force-directed layout in Cytoscape. To study the topology of the built networks, network statistical analysis such as the functional relationship between the numbers of nodes and their degree of connectivity (degree node distribution) was performed utilizing the Cytoscape plug-in Network Analysis (Cline et al., 2007). A number of node and edge attributes were computed, the degree-node-distribution representing the relationship between one particular degree of connectivity and the number of corresponding nodes was visualized as a scatter plot.

3.14 Candidate gene prioritization approach

Reverse engineered networks can predict novel functions for uncharacterized genes but also potential functional associations among known genes. Predictions made by such networks can also be supplemented with prior knowledge or additional data sources to generate new hypotheses for further investigation. Therefore, the network generated by ARACNe or TINGe was complemented with prior knowledge by applying the following strategy. The top 10% highly connected hub genes with a degree of connectivity higher than 30 were prioritized by association with macrophage lineage and activation information using the transcription factors PU.1 and RUNX1 as bait genes. Both have crucial roles as macrophage lineage and activation factors. Using these genes, similarity profiling, data fusion and network-based strategies were performed by applying two prioritization tools, ToppGene (Chen et al., 2009)

and Endeavour (Tranchevent et al., 2008). The results of the different approaches were subsequently combined by the Borda Count method, which was implemented by a Perl script (see **Appendix A Script 1**).

3.15 Common transcription factor binding site prediction

TF binding prediction was performed using the Genomatix Suite (http://www.genomatix.de/). First, promoter models for a set of genes of interest were compiled using the Gene2Promoter module in Genomatix. Subsequently, the Genomatix module for the search of common TF binding sites was applied to determine overrepresented TF binding sites. Significance measure for each TF family is represented by z-score, calculated with a continuity correction using the formula z = (x-E-0.5)/S, where x is the number of found matches in the input data, E is the expected value and S is the standard deviation. A z-score below -2 or above 2 can be considered to be statistically significant. The z-score was subsequently converted as a normal distribution to the corresponding p-value using pnorm command in R.

3.16 Deconvolution analysis by CIBERSORT

CIBERSORT (https://cibersort.stanford.edu/) is an analytical tool for characterizing cell composition of complex tissues or mixed cell populations from their gene expression profiles (Newman et al., 2015). To access the abundance of cell types from hematopoietic origin within complex tissues, the authors first designed and validated a leukocyte signature matrix, which contains 547 genes that distinguish 22 human hematopoietic cell phenotypes including T cells, B cells, plasma cells, NK cells and myeloid compartments. The defined human leukocyte signature was termed “LM22”. To deconvolute the mixture within the input expression matrix, CIBERSORT employs linear support vector regression, a robust machine learning approach, to perform feature selection from the defined cell type signature matrix. The gene expression data generated from PBMC were grouped into 10 groups (Bbu, LPS, Mtu, Cal and Control at both 4 hours and 24 hours) by computing the mean expression over the individual samples. Utilizing this mixture of expression matrix as input, CIBERSORT was applied to run 100 permutations for statistical significance testing. The relative fractions of 18 leukocyte types from LM22 and global p-values for the deconvolution were calculated.

Custom signature gene files can be defined by giving reference sample files for transcriptome profiles from distinct phenotypes and the corresponding phenotype classes file to CIBERSORT. To create human macrophage activation signatures, the expression matrix of 29 human *in vitro* macrophage activation conditions (184 samples including M^b at 0 hour and 28 activated conditions at 24 or 72 hour time points) from **Chapter 4** (Xue et al., 2014) was used as input. After generation of macrophage activation signatures, CIBERSORT calculations were performed on the mixture file derived from the PBMC data (see above).

3.17 Circos plots of stimulation-specific effector molecules

To identify stimulation-specific biomarkers for bacterial and fungal infections, 1,427 surface markers (de Souza et al., 2012) and 55 cytokines including chemokines (Costantini et al., 2009) that are present in 12,029 pre-filtered genes were extracted from the PBMC data. The markers were tested if FC > 1.5 and FDR adjusted p-value < 0.05 when comparing each condition against the corresponding control: if yes, one link between the TF and the condition was made; otherwise, no link between them would be generated. The Circos program (Krzywinski et al., 2009) was applied to visualize the different diagrams for surface markers and cytokines. Before running Circos, a few specific input files such as “karyotype”, “histogram” and “link” with a certain format were generated in batch by using a Perl script (see **Appendix A Script 2**). For each plot, one complete circle was split into 9 parts as virtual “chromosomes”, one for all the markers and the other 8 representing 8 conditions. The length of each “chromosome” reflecting the number of involved markers that were changed for the corresponding condition. The links were colored according to the conditions. The configuration file of one Circos run for the surfaceome plot (**Figure 6.4.8**) is provided in **Appendix A Script 3**.

3.18 K-means clustering

To identify gene clusters with similar expression patterns, K-means clustering algorithm (partitioning clustering) in PGS was applied. The preprocessed murine tissue macrophage data as described in **Chapter 3.2.2** from 17 selected samples as embryonic development-associated groups including EMP, A1, A2, embryonic F4/80^{high} kidney cells, Kupffer cells and microglia (abbreviations see **Table 7.1**) were extracted and the mean expression values for each group were computed. The City block distance function was employed and the Davies-Bouldin index (Davies and Bouldin, 1979) was calculated before clustering to check the

optimal number of clusters. As a result, the 17,888 present genes were clustered into 36 clusters.

3.19 Wet lab validation methods

Molecular and cellular experimental validations on different levels, such as reverse transcription polymerase chain reaction (RT-PCR), western blotting and enzyme-linked immunosorbent assay (ELISA), were performed via collaboration with other lab members. These data are only included in this thesis to illustrate the validity of the computational models derived within this thesis. The respective collaborators who performed the experiments are mentioned throughout the text.

4 Transcriptome-based network analysis of human macrophage from diverse signal inputs

Most of the work described in **Chapter 4** was published in Immunity (Xue et al., 2014).

4.1 Extension of M1 versus M2 polarization to a multi-dimensional model

Within the Department of Genomics and Immunoregulation at the LIMES-Institute, human peripheral blood mononuclear cells (PBMC) were isolated and $CD14^+$ monocytes were derived by magnetic-activated cell sorting (MACS). Macrophages were generated from $CD14^+$ monocytes by differentiation with GM-CSF or M-CSF, and as control cells dendritic cells (DCs), T cells, B cells, and Natural Killer (NK) cells were also isolated or generated and used for genome-wide transcriptional profiling (**Figure 4.1.1**, **Table 4.1**). To assess overall sample-to-sample relationships, macrophages (baseline, M^b) were clearly distinct from other cell types including DCs in co-regulation analysis as it is shown in **Figure 4.1.2.** The difference between M^b and DCs on protein level was then confirmed by flow cytometry experiments, which were performed by my collaborators in the department (mainly Susanne V. Schmidt and Heidi Theis) (**Figure 4.1.3**).

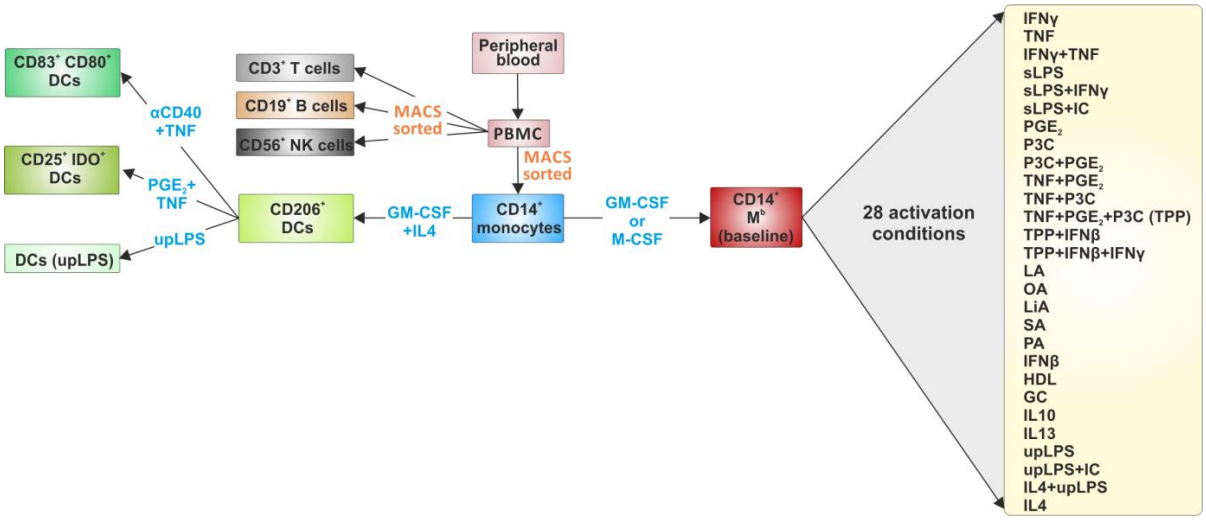

Figure 4.1.1 Schema of isolation and generation of cells used in this dataset. Figure was adapted from (Xue et al., 2014). M-CSF: macrophage colony-stimulating factor; GM-CSF: granulocyte-macrophage colony-stimulating factor; GC: glucocorticoid; IC: immune complexes; PGE_2: prostaglandin E2; P3C: Pam3CysSerLys4; TPP: TNF+PGE_2+P3C; PA: palmitic acid; OA: oleic acid; LA: lauric acid; LiA: linoleic acid; SA: stearic acid; sLPS: standard lipopolysaccharide; upLPS: ultrapure lipopolysaccharide; HDL: high density lipoprotein.

Table 4.1 Gene expression profiling data sample overview

Cell type	Count of array samples
B cell	3
Dendritic cell	33
Macrophage	299
baseline	53
GC	3
HDL	4
IFNβ	3
IFNγ	38
IFNγ_TNF	3
IL10	3
IL13	3
IL4	40
IL4_upLPS	6
LA	6
LiA	6
M1/2	2
OA	12
P3C	3
P3C_PGE_2	3
PA	12
PGE_2	3
SA	6
sLPS	8
sLPS_IC	3
sLPS_IFNγ	3
TNF	6
TNF_P3C	3
TNF_PGE_2	3
TPP	46
TPP_IFNβ	3
TPP_IFNβ_IFNγ	3
upLPS	9
upLPS_IC	3
Monocyte	15
NK cell	3
T cell	31
Grand Total	384

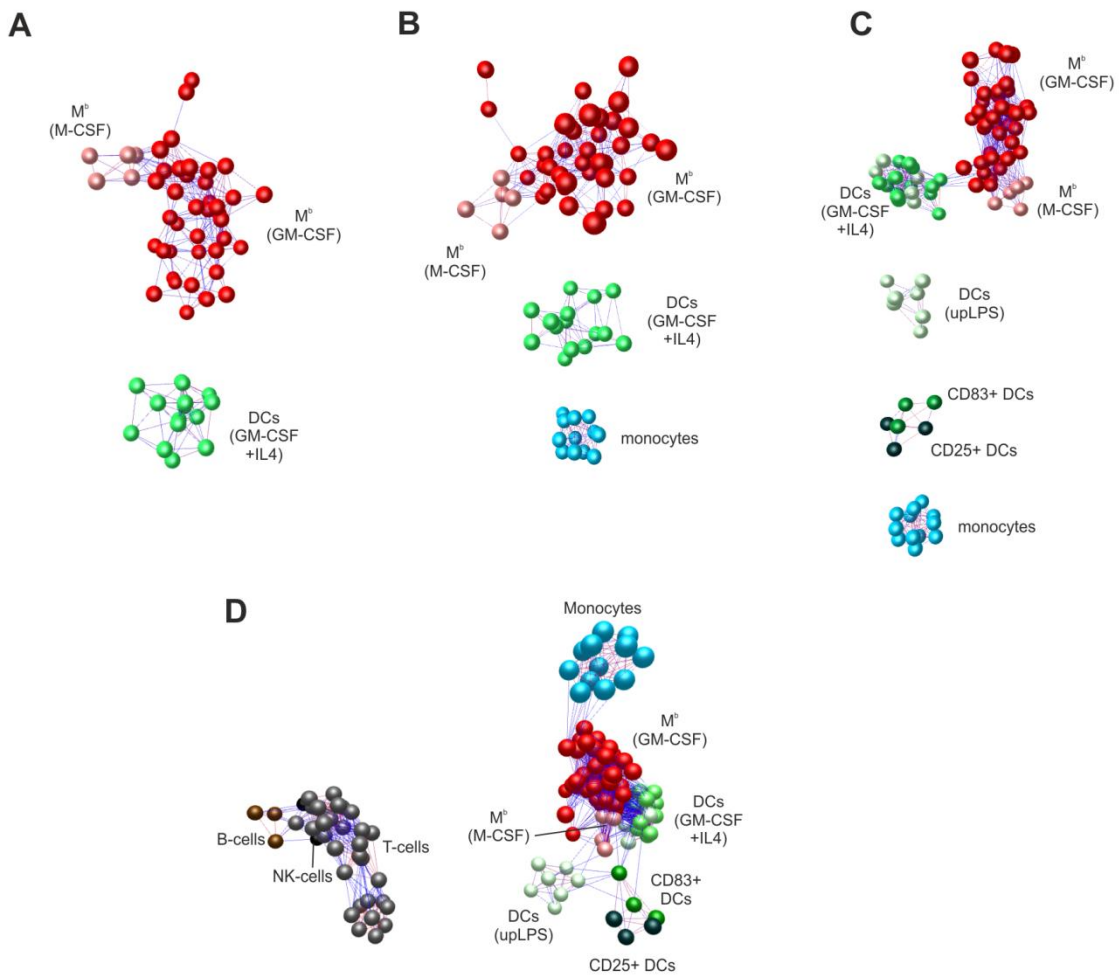

Figure 4.1.2 Relationship between macrophages and other immune cells. Figure was adapted from (Xue et al., 2014). **A** Sample-sample co-regulation network (visualized in 3-dimensional space) of monocyte-derived macrophages (baseline, M^b) induced by M-CSF or GM-CSF with monocyte-derived cells (DCs) induced by GM-CSF+IL4. Sample-sample co-regulation networks additionally including **B** monocytes, **C** $CD83^+$ DCs, $CD25^+$ DCs and upLPS-stimulated DCs, **D** T cells, B cells and NK cells.

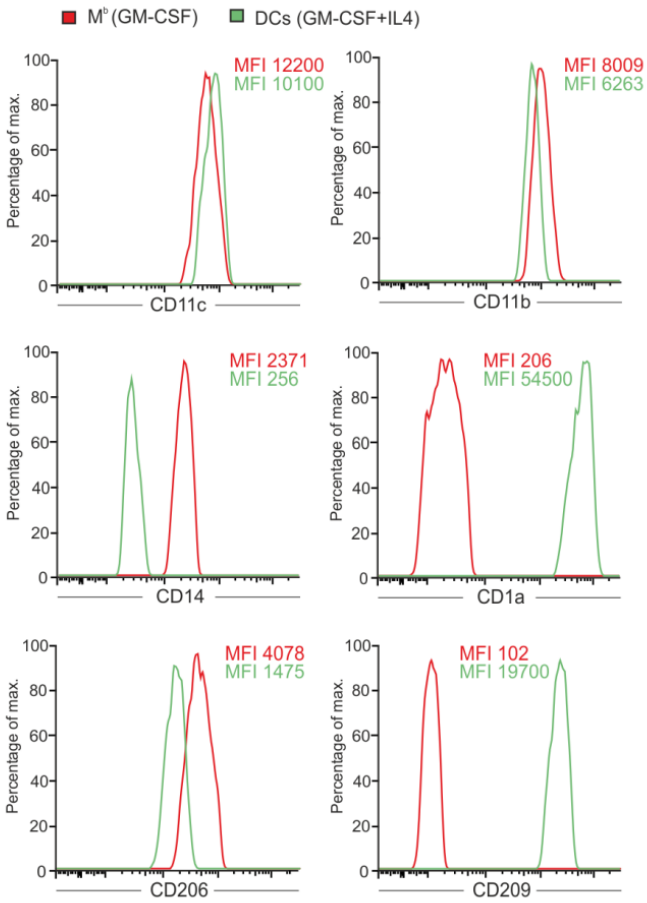

Figure 4.1.3 Comparison of surface marker expression for M^b and DCs. Figure was adapted from (Xue et al., 2014). Representative histograms of expression of cell surface molecules CD11c, CD11b, CD14, CD1a, CD206 and CD209 on M^b (GM-CSF, red) and DCs (GM-CSF+IL4, green). Data were generated by Susanne V. Schmidt and Heidi Theis.

The monocyte-derived macrophages (M^b) were stimulated with 28 different stimuli including pattern recognition receptor ligands, cytokines and metabolic cues for a certain time period (ranging from 0.5 to 72 hours) as shown in **Figure 4.1.1**. To better understand the complexity of transcriptional regulation after macrophage activation, the complete transcriptomes of M^b and macrophages activated by 28 stimulation conditions were analyzed. To determine the overall relationship of these activation states, co-regulation analysis was applied as it is shown in **Figure 4.1.4**. In agreement with the existing dichotomous model, a virtual axis was formed, where macrophages at M^b were localized in between macrophages stimulated with IFNγ (so-called M1, $M^{IFN\gamma}$) and IL4 (so-called M2, M^{IL4}) (**Figure 4.1.4A**). Adding other conditions linked to M1 (sLPS, TNF) or M2 (IL13) polarization (Biswas and Mantovani, 2010)(**Figure 4.1.4B**) did not change the overall M1 and M2 axis (**Figure 4.1.4A**). Including further M1- and M2-associated stimuli (IFNγ+TNF, IL10) increased the variance in the

correlation matrix but the overall bipolar structure maintained (**Figure 4.1.4C**). However, when adding stimuli that have not been linked to either M1 or M2 polarization, such as free fatty acids, high density lipoprotein (HDL), or combinations of stimuli associated with chronic inflammation (TNF+PGE_2+P3C, TPP), the co-regulation network revealed further extensions that broke the initial bipolar 'axis' structure (**Figure 4.1.4D**). When all 29 conditions were included in the co-regulation network, the multi-dimensional structure of macrophage activation was formed and expanded (**Figure 4.1.4E**). Furthermore, when adding samples generated at earlier time points after stimulation, the multi-dimensionality of macrophage activation was shown to consist of a dense network of individual signatures (**Figure 4.1.4F**).

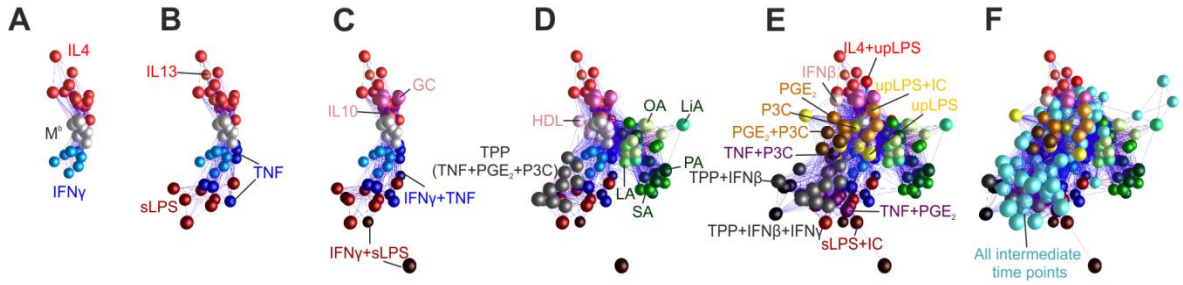

Figure 4.1.4 Co-regulation networks of 297 macrophage transcriptomes representing 29 conditions. Figure was adapted from (Xue et al., 2014). Each node represents one sample generated at end of activation time points (**A-E**) or also including data at intermediate time points (**F**).

These findings were validated by using several bioinformatic approaches such as Self-Organizing-Map (SOM) clustering and Pearson correlation coefficient matrix (PCCM). Samples belonging to one condition have been grouped into one eigenvalue (mean expression over samples). Performing SOM clustering on the 17 conditions shown by co-regulation analysis in **Figure 4.1.4D** revealed that every stimulus was characterized by a specific cluster structure (**Figure 4.1.5**), which further supported the multi-dimensional model. Similarly, a bipolar structure was not identified within the PCCM, but rather a condition-specific multi-dimensional structure of correlation coefficients in 5 major clusters, which could be further divided into 10 clusters as shown in **Figure 4.1.6**. PCCM results were quite similar when using the complete transcriptome or only 1,000 most variable probes.

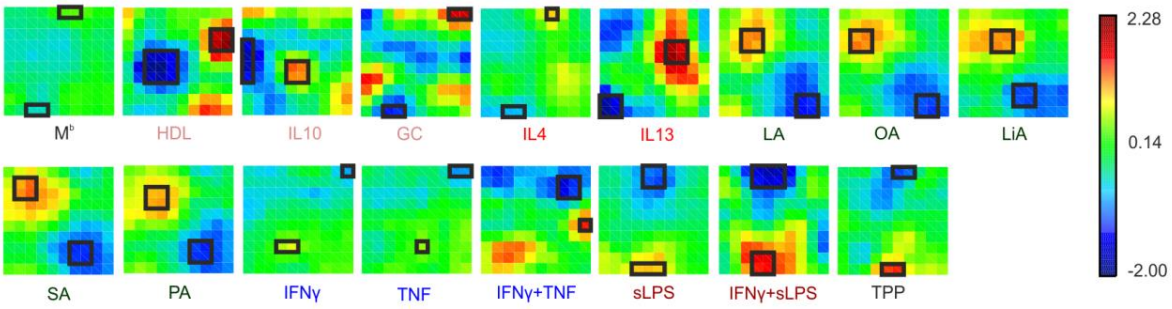

Figure 4.1.5 Self-Organizing Map (SOM) clustering using samples displayed in Figure 4.1.4D (17 conditions). Figure was adapted from (Xue et al., 2014). Each transcriptome is clustered into 10 x 10 clusters. Each square represents the z-transformed expression value (z-score) as eigenvalue for cluster genes as a heatmap. Clusters with the top up- or down-regulated genes for each stimulus are marked with a frame.

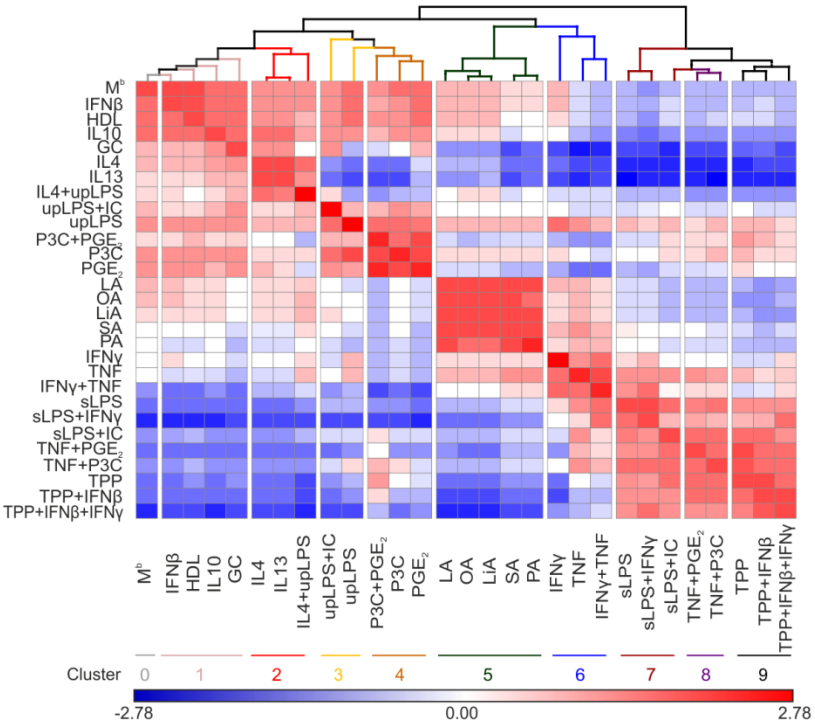

Figure 4.1.6 Heatmap of hierarchically clustered Pearson correlation coefficient matrix (PCCM). Figure was adapted from (Xue et al., 2014). PCCM was calculated based on 1,000 most variable probes for 29 conditions. Pearson correlation coefficients are standardized from -2.78 to 2.78 (blue via white to red).

Furthermore, sum vectors in three-dimensional space were built by using the coordinates of the samples defined by co-regulation analysis within the 10 clusters defined by PCCM. As a result, a model of macrophage activation best described by multi-dimensionality of transcriptional programs was proposed as illustrated in **Figure 4.1.7**.

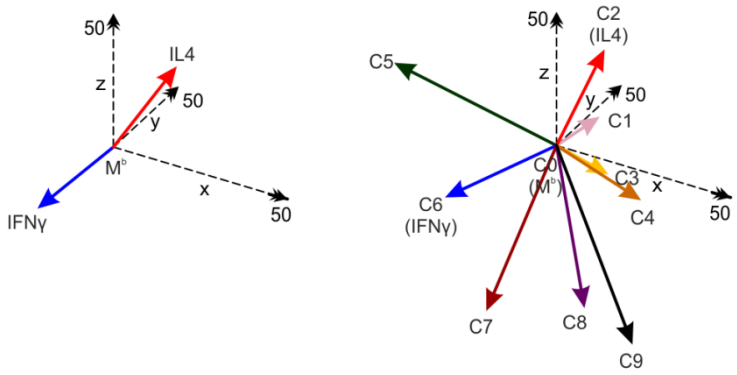

Figure 4.1.7 Multi-dimensional model (3D) based on the 10 clusters. Figure was adapted from (Xue et al., 2014). Sample values (coordinates) defined by co-regulation network (from **Figure 4.1.4E**). Baseline macrophages (M^b) are set as origin, 9 activation states are represented by colored arrows, x-, y- and z-axes are in dashed lines with double arrows.

Taken together, these data clearly extend the current model of M1 versus M2 polarization to a multi-dimensional model of macrophage activation.

4.2 Identification of genes specifically associated with distinct stimuli

To determine whether the different stimuli could be distinguished on the gene level within the complete multi-dimensional model, 9,498 genes, which were expressed in at least one condition from all macrophage samples at end of activation states, were kept for further analysis. SOM clustering was performed for these genes so that the z-transformed expression values (z-scores), which enable the selection of genes specifically regulated and enriched for individual stimuli, were calculated. As a result, genes that were selectively elevated in only one of the stimulation conditions were identified. For instance, glucocorticoid (GC) selectively induced SESN1, while palmitic acid (PA) selectively increased the expression of HES4 (**Figure 4.2.1**-**4.2.2**). However, there were also conditions, for which a single gene was insufficient to distinguish between closely associated stimuli, e.g. SERINC2 was induced by PGE_2, but also by PGE_2+P3C (**Figure 4.2.2**), indicating that gene combinations are necessary to discriminate complex input signals on transcriptional level. In general, although some activation stimuli might be associated with the induction of single genes, the assessment of a substantial number of markers as representatives for different activation programs of macrophages will be still important in future studies.

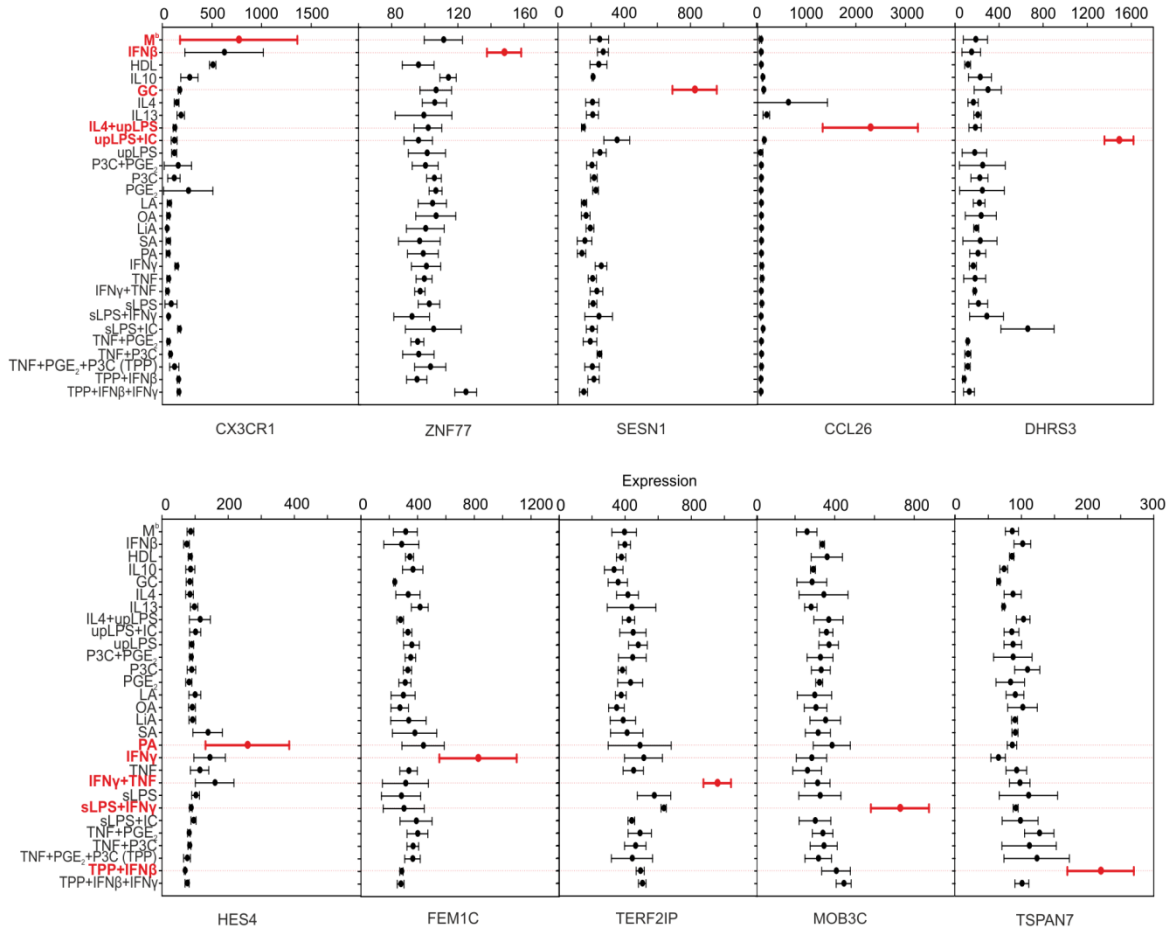

Figure 4.2.1 Genes with selective expression associated with distinct stimuli. Figure was adapted from (Xue et al., 2014). Absolute expression values (mean ± SD) of genes defined by SOM clustering to be highly expressed for a particular stimulation condition. Shown here are genes enriched in M^b (baseline), IFNβ, GC, IL4+uLPS, upLPS+IC, PA, IFNγ, IFNγ+TNF, sLPS+IFNγ, or TPP+IFNβ.

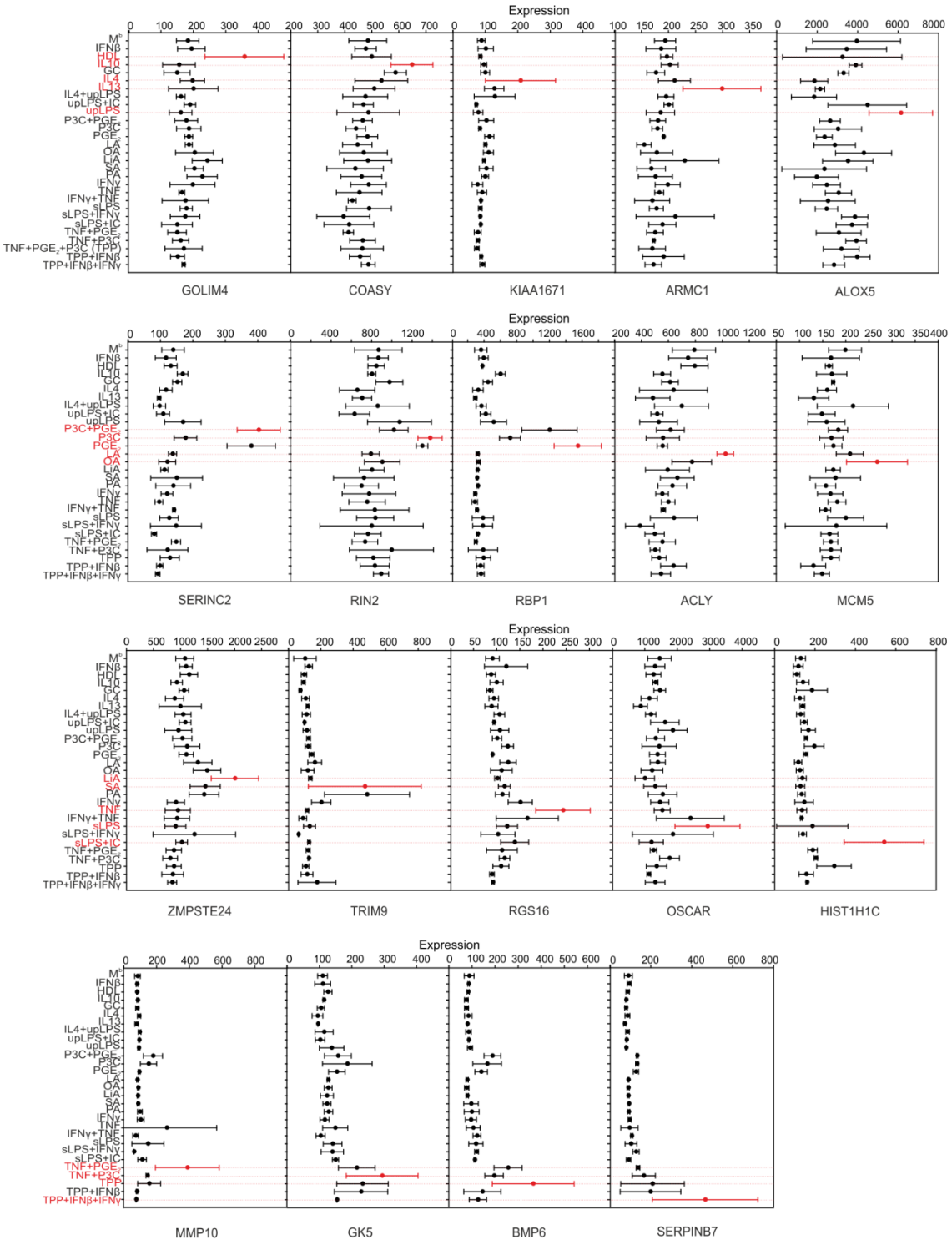

Figure 4.2.2 Genes with selective expression associated with distinct stimuli (continued). Figure was adapted from (Xue et al., 2014). Absolute expression values (mean ± SD) of genes defined by SOM clustering to be highly expressed for a particular stimulation condition. Shown are the genes induced by the conditions not shown in **Figure 4.2.1**.

4.3 Network analysis identifies stimulus-associated transcriptional programs of macrophage activation

To further investigate input signal-specific gene sets that are associated with the corresponding macrophage activation states, weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) was applied on 29 conditions. It identifies pairwise Pearson correlation across genes, and then classifies the complete transcriptome into a number of gene modules. The WGCNA analysis was mainly performed by my collaborators in the department (mainly Thomas Ulas). On the basis of 9,498 present genes as described above, 49 distinct co-expression modules containing 27 to 884 genes per module were identified. To determine the module eigengene (ME), which is the first principal component of the respective module, the expression data from different genes within each computed module were used. Afterwards, these 49 MEs were correlated to the 29 experimental conditions. As examples the eigengene expression of modules 8, 15, and 30 (**Figure 4.3.1**), which are highly enriched in IFNγ, IL4 and TPP stimulation conditions, were visualized as colored bar plots. The resulting ME-to-condition correlation matrix was then visualized as a heatmap (**Figure 4.3.2**). The $M^{IFN\gamma}$ and M^{IL4} showed prominent ME patterns, while other stimuli clearly displayed divergent patterns further supporting a multi-dimensional model of macrophage activation. For example, stimulation with TPP (M^{TPP}) induced a strong signal in modules 30, 32, and 33, which could not be observed in $M^{IFN\gamma}$, or M^{IL4}. Combined TNF, PGE_2 and TLR ligand P3C activation was introduced as a reductionist model of chronic granulomatous inflammation such as in tuberculosis, or granulomatous listeriosis (Marino et al., 1997; Popov et al., 2006; Shay and Simon, 2012). It has been demonstrated that genes like IL10, CD25, COX-2, and indoleamine 2,3-dioxygenase (IDO) are expressed in macrophages in human granulomatous structures and are induced in human macrophages after stimulation with the combination of the aforementioned factors (TNF, PGE_2, TLR2 ligand P3C) (Popov et al., 2006; Popov et al., 2008) suggesting that these host factors characterize the transcriptional program during chronic inflammation.

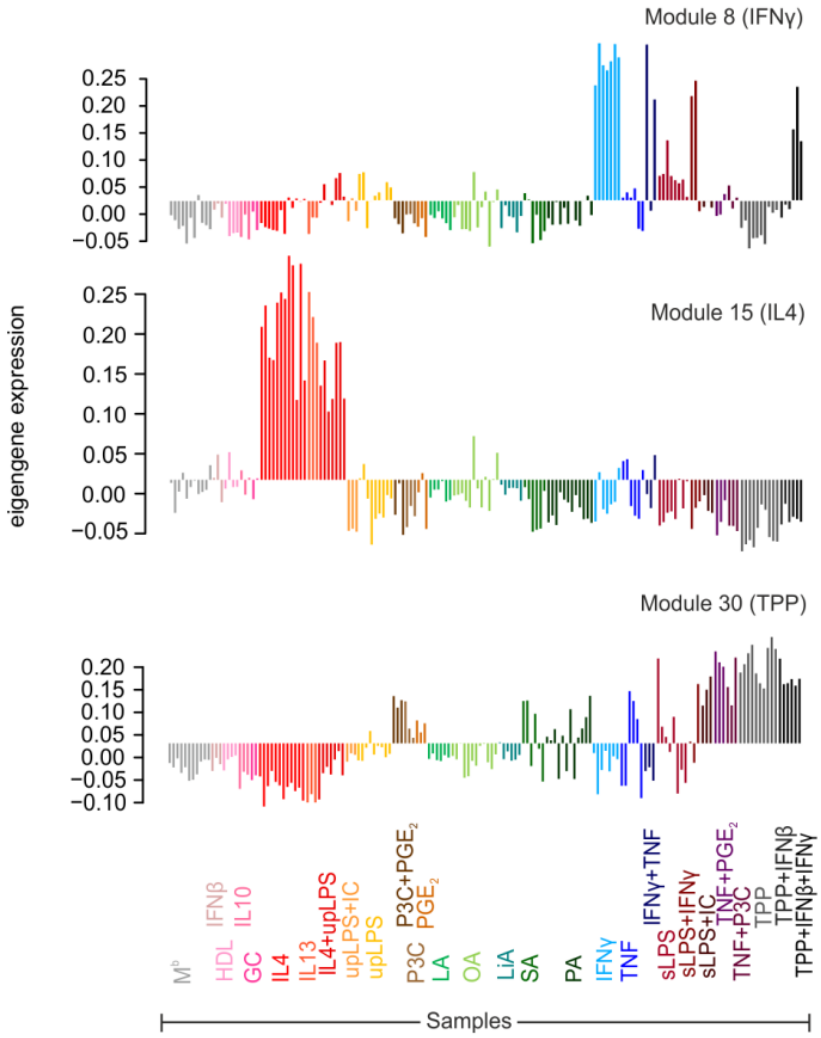

Figure 4.3.1 Bar plot visualization of the eigengene expression of modules 8, 15, and 30 in the 29 stimulation conditions. Figure was adapted from (Xue et al., 2014). The samples colored according to 29 conditions are shown in x axis while eigengene expression in y axis. Data were produced by Thomas Ulas.

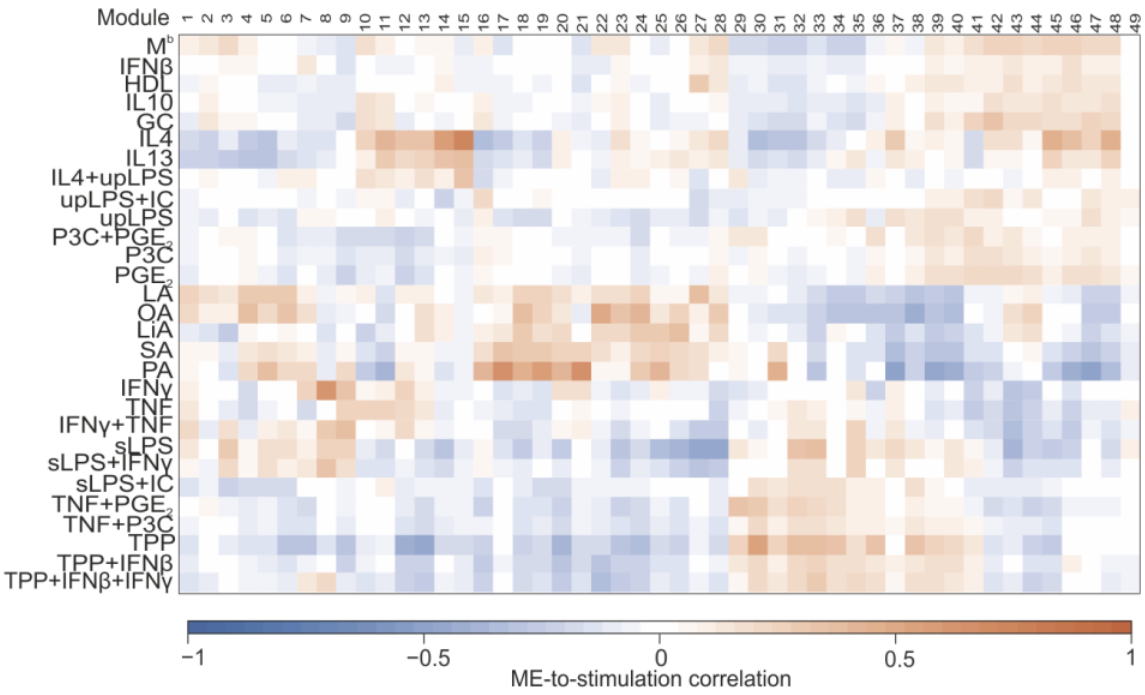

Figure 4.3.2 49 stimulus-specific gene modules identified by weighted gene co-expression network analysis (WGCNA). Figure was adapted from (Xue et al., 2014). Heatmap showing the correlation of the module eigengene (first principal component, ME) to the stimulation conditions. Blue means negative correlation while orange means positive correlation. Data were produced by Thomas Ulas.

Next, the gene modules correlated with the specific stimulation conditions IFNγ, IL4 and TPP were extracted to link respective module genes to biological information provided by the Gene Ontology (GO) consortium (http://geneontology.org/). GO enrichment analysis was performed by the combination of the GO enrichment calculations (BiNGO) and network visualization of enriched GO-terms (EnrichmentMap) (Figure **4.3.3**). This analysis confirmed the major functional differences between $M^{IFN\gamma}$ and M^{IL4}, e.g. that an IFNγ-driven response was associated with the induction of pro-inflammatory genes, whereas these genes were depressed in the IL4-driven response. More importantly, TPP stimulation induced a gene expression pattern associated with chronic inflammation, including GO terms such as 'chronic inflammatory response'. This approach was validated by interrogating these gene sets with GO analysis, pathway analysis and transcription factor (TF) binding prediction tools provided by InnateDB, with the same outcome (data not shown).

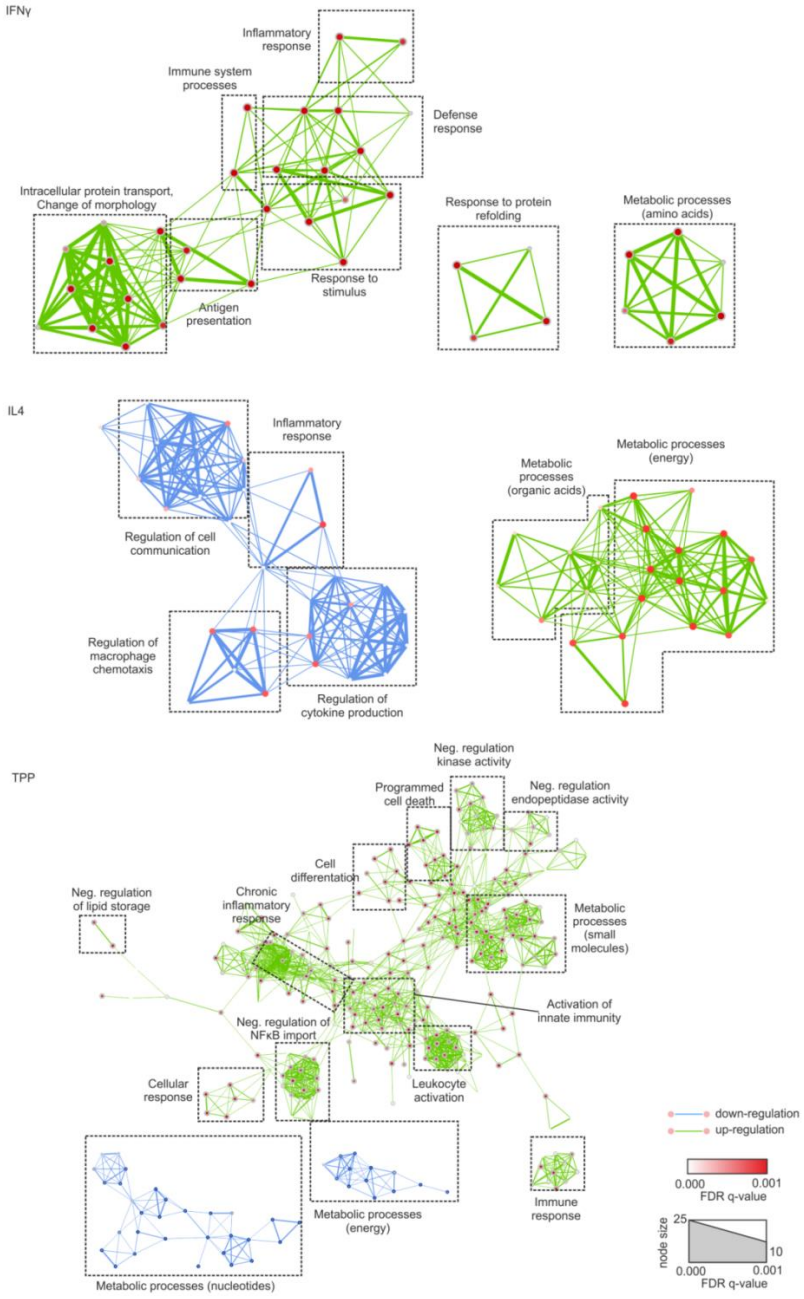

Figure 4.3.3 GO network of IFNγ-, IL4- and TPP-associated modules. Figure was adapted from (Xue et al., 2014). Network visualization of GO enrichment analysis of modules 7-9 (positively correlated) and 43, 44, 48 (negatively correlated) for IFNγ stimulation, 13-15 (positively correlated) and 30, 5, 32 (negatively correlated) for IL4 stimulation, 30, 32-33 (positively correlated) respectively 12, 13, and 20 (negatively correlated) for TPP stimulation. Node size and node color darkness correspond with the enrichment FDR adjusted p-value (q-value) of the GO term. Green edges stand for up-regulation of the involved genes while blue edges represent down-regulation. Edge thickness shows overlap of genes between neighbor nodes. Data were produced by Thomas Ulas.

Subsequently, TFs within the IFNγ, IL4, and TPP-associated modules were determined using the Genomatix database and visualized them as co-regulation networks as shown in **Figure**

4.3.4. This analysis revealed STAT1 as a central hub in $M^{IFN\gamma}$ and STAT6 as a hub in M^{IL4}. Moreover, it has been observed that additional TFs are related to these activation programs such as STAT2, IRF7 and IRF9s for IFNγ activated cells as well as IRF4 and the forkhead box proteins FOXQ1 and FOXD2, which were not previously associated with the IL4 activation. In the case of macrophages stimulated with TPP, the TF co-regulation network also included STAT4 as well as TFs mediating both inflammation and metabolism (HIF1A, (Shay and Simon, 2012)), or TFs associated with negative regulation of TLR signaling (HEY1 (Hu et al., 2008)), or macrophage activation (TGIF1, (Ramsey et al., 2008)). Other TFs identified in this network have not yet been linked to macrophage activation as determined by pubatlas.org-based literature mining. This analysis identified activation-associated gene sets, which are responsible for essential biological functions of various macrophage subtypes. These gene sets assemble not only well-known TFs linked to the main macrophage activation programs, but also TFs that have not yet been associated with any activation states. Taken together, this large dataset of macrophage activation enables the establishment of stimulus-specific transcriptional networks in human macrophages.

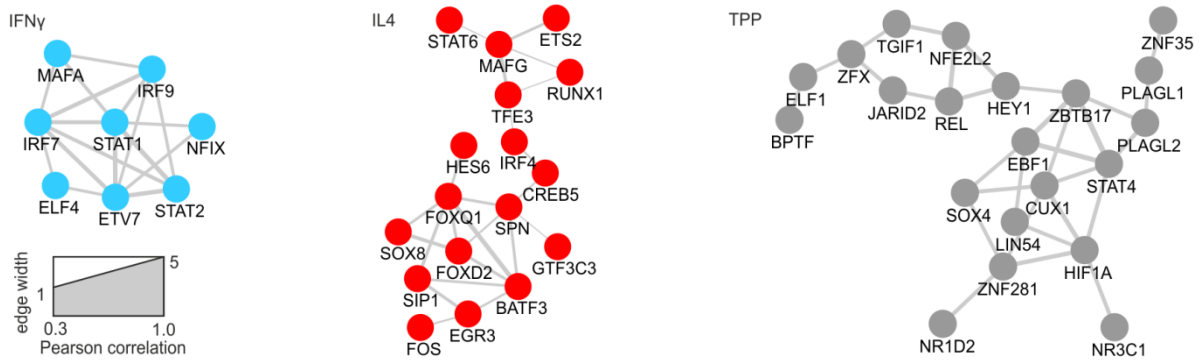

Figure 4.3.4 Co-regulation network of module-specific transcription factors (TFs). Figure was adapted from (Xue et al., 2014). IFNγ: modules 7-9; IL4: modules 13-15; TPP: module 30, 32-33. Edge width shows the Pearson correlation between each TF pair.

4.4 Novel phenotypes and function of macrophages activated by TNF, PGE_2, and TLR2 ligand

To demonstrate phenotypic and functional differences of M^{TPP} to the prototype $M^{IFN\gamma}$ and M^{IL4}, experimental validation on molecular and cellular levels was performed by my collaborators in the department (mainly Susanne V. Schmidt, Heidi Theis, Astrid Draffehn and Andrea Nino-Castro). As it has been determined by PCCM (**Figure 4.1.6**) and WGCNA (**Figure 4.3.2**), M^{TPP} have a transcriptomic signature distinct from $M^{IFN\gamma}$ or M^{IL4} macrophages.

Within differentially expressed genes (FC > 2, FDR adjusted p-value < 0.05) between M^{TPP}, $M^{IFN\gamma}$, M^{IL4}, and M^{b}, a total number of 51 cell surface markers expressed selectively on M^{TPP} was identified, i.e. they were only induced in M^{TPP} but not in $M^{IFN\gamma}$, M^{IL4} or M^{b}. Significantly high expression for CD14, CD23, CD25, CXCR7 and CD197 on M^{TPP} (p-value < 0.05) was verified by flow cytometry, while CD86 was elevated on both $M^{IFN\gamma}$ and M^{TPP} (**Figure 4.4.1**).

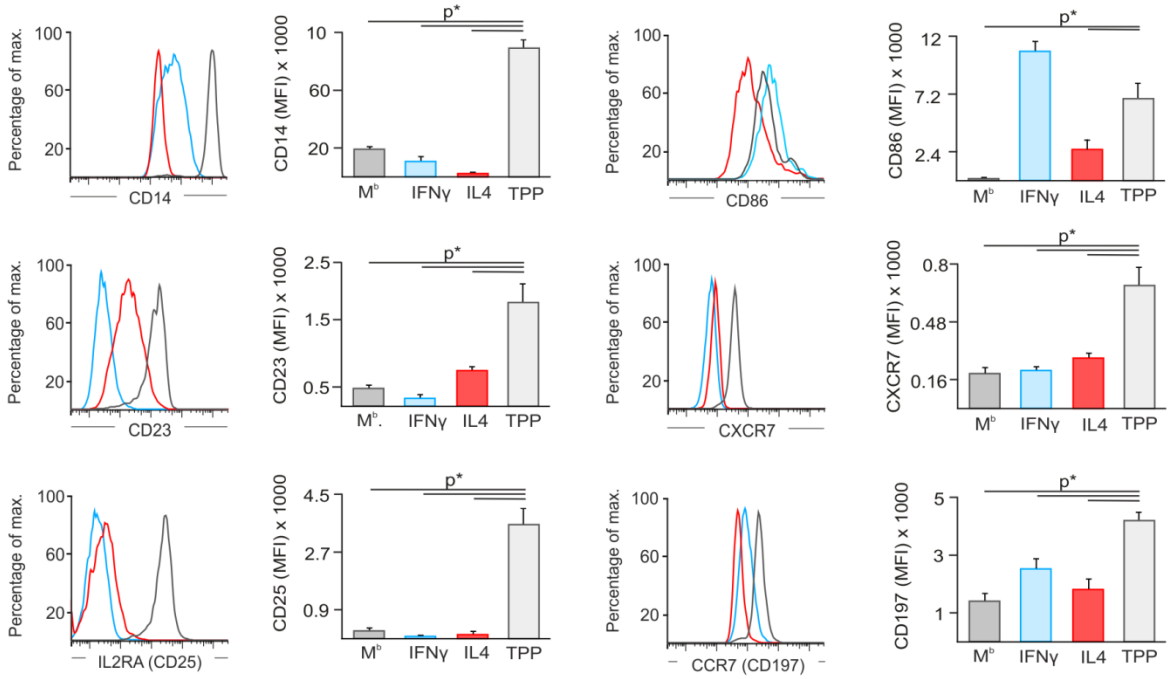

Figure 4.4.1 Flow cytometry of CD14, CD23, CD25, CD86, CXCR7 and CD197 in macrophages. Figure was adapted from (Xue et al., 2014). Four types of macrophages are marked in different colors: M^{b} in dark grey, IFNγ in light blue, IL4 in red and TPP in light grey. Mean fluorescence intensities (MFI) of at least three independent experiments (mean and s.e.m.; p* < 0.05 Student's t-test). Data were generated by Susanne V. Schmidt.

Besides surface molecules, a set of TFs was found to be induced in M^{TPP} but not in $M^{IFN\gamma}$ or M^{IL4}: STAT4 as an example. It was confirmed by western blotting that STAT4 protein expression is only induced in M^{TPP} (**Figure 4.4.2**).

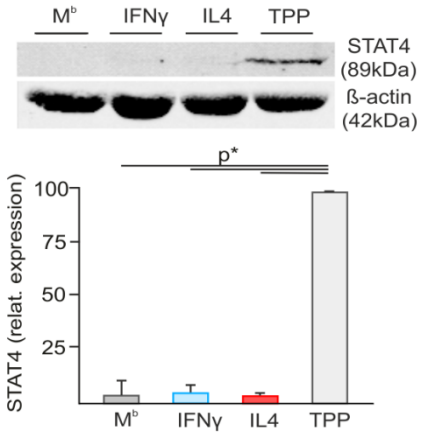

Figure 4.4.2 Western blot analysis of STAT4 and β-actin. Figure was adapted from (Xue et al., 2014). STAT4 expression was normalized to ß-actin expression and set to 100% in M^{TPP} (TPP) (mean and s.e.m.; p* < 0.05 Student's t-test). Data were generated by Astrid Draffehn and Heidi Theis.

Moreover, further differences between M^{TPP} and $M^{IFNγ}$ or M^{IL4} were observed on cytokine secretion level. For instance, as shown in **Figure 4.4.3**, CXCL5 secretion was significantly induced by M^{TPP} and to a lesser extent by $M^{IFNγ}$ (p-value < 0.05), but not by M^{IL4}, while IL1α was only secreted by M^{TPP}.

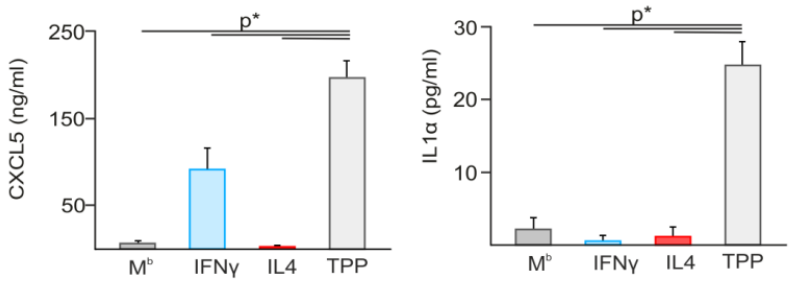

Figure 4.4.3 Cytokine secretion in macrophages. Figure was adapted from (Xue et al., 2014). CXCL5 and IL1α levels in supernatants of macrophage cultures: M^b (dark grey), IFNγ (light blue), IL4 (red) and TPP (light grey) activated macrophages (mean and s.e.m.; p* < 0.05 Student's t-test). Data were generated by Susanne V. Schmidt and Heidi Theis.

Since macrophage activation has impact on T cell proliferation, T cell activation by CD3 and CD28 beads in presence or absence of macrophages was assessed by CFSE dilution. As a result, T cell proliferation was decreased by M^b although not statistically significant (p-value < 0.05), and there was no difference in CD3 and CD28 stimulated T cell proliferation in the presence of $M^{IFNγ}$ and M^{IL4} (**Figure 4.4.4**). M^{TPP}, however, strongly repressed T cell

proliferation, which indicates that macrophage activation by TPP induced a distinct functional program from IFNγ- or IL4-activated macrophages.

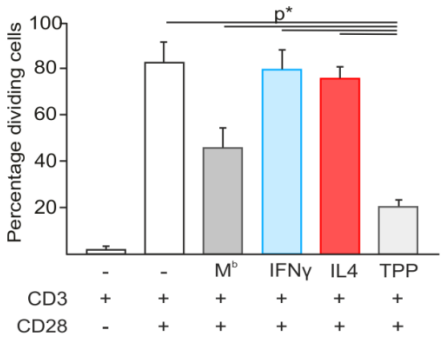

Figure 4.4.4 T cell activation in presence or absence of macrophages. Figure was adapted from (Xue et al., 2014). The percentage of dividing CD3 and CD28 stimulated T cells in presence or absence of M^{b} (dark grey), IFNγ (light blue), IL4 (red) and TPP (light grey) activated macrophages (mean and s.e.m.; p* < 0.05 Student's t-test). Data were generated by Andrea Nino-Castro.

MicroRNAs (miRNAs) are small RNAs that regulate gene expression after mRNA transcription. Therefore, miRNA expression in IFNγ, IL4, and TPP conditions was assessed globally by miRNA-Seq to detect their difference on post-transcriptional level. And the analysis of the miRNA-Seq data was performed by my collaborators Martina Emde and Jil Sander. As it is shown in **Figure 4.4.5**, M^{TPP} clearly differed from $M^{IFNγ}$ and M^{IL4} at the miRNA level: on the one hand, M^{TPP} had prominent hsa-miR-125a-5p expression and a lack of $M^{IFNγ}$- (hsa-miR-23b-3p) or M^{IL4}-associated miRNAs (e.g. hsa-miR-125b-5p, hsa-miR-99a-5p); On the other hand, a set of miRNAs had significant decrease of expression in M^{TPP} when comparing to IFNγ and IL4 activated cells (FDR adjusted p-value < 0.05).

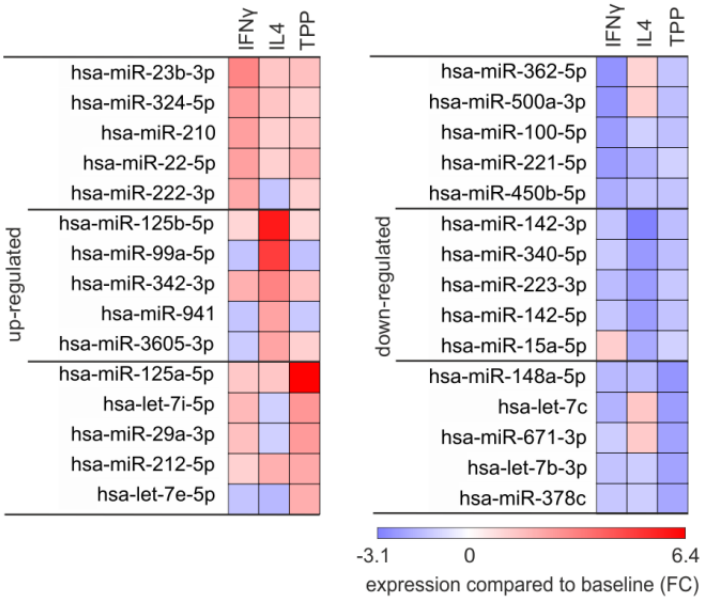

Figure 4.4.5 Differential expression of miRNAs between IFNγ-, IL4- and TPP-induced macrophage activation. Figure was adapted from (Xue et al., 2014). Heatmap showing fold changes of highly abundant miRNAs up- or down-regulated (FC > 2, FDR adjusted p-value < 0.05) in M1 (IFNγ) or M2 (IL4), or M^{TPP} (TPP) compared to M^{b} (baseline). Fold changes colored from blue via white to red. Data were generated by Martina Emde and Jil Sander.

Collectively, macrophages that are different in their global transcriptional program from $M^{IFNγ}$ or M^{IL4}, such as M^{TPP} macrophages, are also distinct in their phenotype and function. These findings further support the multi-dimensional model of macrophage activation.

4.5 Overlay *in vitro* expression data to human alveolar macrophages

Next, the relevance of the *in vitro* macrophage activation model for *in vivo* macrophage biology was evaluated. To address whether particular activation programs such as those described in **Chapter 4.3** can be found in human tissue macrophages, two datasets of human alveolar macrophages obtained by bronchoalveolar lavage (Shaykhiev et al., 2009; Woodruff et al., 2005) consisting of samples from non-smokers, smokers, and COPD patients were collected from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), and were integrated into one. After normalization and filtering steps, co-regulation analysis of samples was performed on the basis of differentially expressed genes among clinical groups (FC > 2.0, FDR adjusted p–value < 0.05). Three major clusters reflecting the three patient groups were revealed by co-regulation analysis (**Figure 4.5.1**) supporting distinct transcriptional programs in macrophages from the three groups.

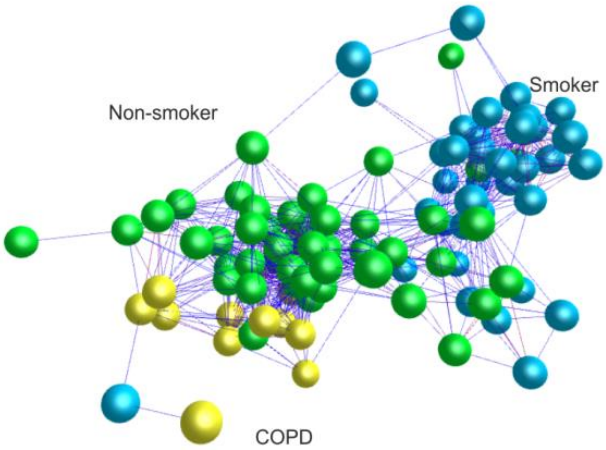

Figure 4.5.1 Relationships throughout samples from alveolar macrophages. Figure was adapted from (Xue et al., 2014). Co-regulation network of human alveolar macrophages (n = 100) from two studies (Shaykhiev et al., 2009; Woodruff et al., 2005) using 374 differentially expressed genes between non-smokers (n = 39) and smokers (n = 49) or COPD (n = 12) patients (|FC| > 2.0, FDR adjusted p–value < 0.05).

As a next step, stimulus-specific gene modules identified by WGCNA were utilized as 49 gene sets from *in vitro* conditions for Gene Set Enrichment Analysis (GSEA). As positive controls, GSEA was first applied to the comparison of IFNγ-, IL4-, TPP- and palmitic acid (PA)-stimulated macrophages (M^{PA}) with baseline macrophages. Normalized enrichment scores (NES) as well as enrichment p-values for 49 gene sets were calculated and visualized as a Volcano plot (**Figure 4.5.2**). As expected, only those gene sets or gene modules, which were most highly correlated with $M^{IFN\gamma}$, M^{IL4}, M^{TPP} or M^{PA} in the WGCNA analysis, showed the highest positive NES and lowest significant p-values for the corresponding stimuli.

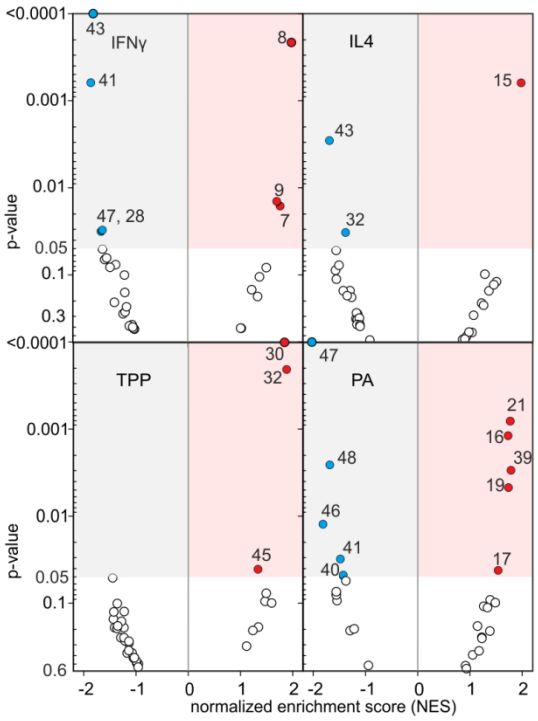

Figure 4.5.2 Results of gene set enrichment analysis (GSEA) on 4 conditions as positive controls. Figure was adapted from (Xue et al., 2014). Volcano plots of normalized enrichment scores (NES) and enrichment p-values based on GSEA for the stimuli IFNγ, IL4, TPP, and PA. Red circles: gene sets positively significantly enriched (NES > 1, p-value < 0.05); blue circles: gene sets significantly depleted (NES < -1, p-value < 0.05).

Since the positive controls of this approach worked properly, GSEA was subsequently applied to the clinical groups, smokers and chronic obstructive pulmonary disease (COPD) patients, in comparison to non-smokers (**Figure 4.5.3**). Surprisingly, for smokers a glucocorticoid-(GC) associated gene module (41, WGCNA) was most significantly enriched (p-value < 0.05) followed by several gene modules associated with free fatty acids and also IL4 and TPP stimulation. This indicated that there is a complex network of stimuli and signaling pathways playing a role in alveolar macrophages in smokers. Previous literature described the presence of M2-like signatures in COPD patients (Shaykhiev et al., 2009). However, no enrichment of IL4- or IL13-associated signatures was observed in COPD patients by using this data driven approach. Furthermore, the promising signature of enriched modules, which was observed in smokers, was completely lost in COPD patients. In the meanwhile, the most significantly down-regulated (NES < -1, p-value < 0.05) WGCNA module in COPD patients was module 8 (associated with IFNγ stimulation), which was also significantly decreased in smokers.

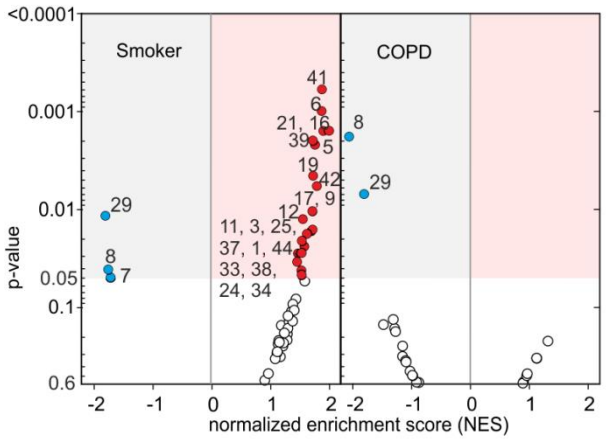

Figure 4.5.3 Results of gene set enrichment analysis (GSEA) on patient sample groups. Figure was adapted from (Xue et al., 2014). Volcano plots of normalized enrichment scores (NES) and enrichment p-values of the same gene sets applied to data from alveolar macrophages derived from smokers and COPD patients. Red circles: gene sets positively significantly enriched (NES > 1, p-value < 0.05); blue circles: gene sets significantly depleted (NES < -1, p-value < 0.05).

To further look into the profound transcriptional signatures, positively enriched modules and negatively enriched modules (p-value < 0.01) for smokers and COPD patients, respectively, were taken for GO enrichment analysis. As a result, network visualization of GO enrichment analysis on smoker related modules results in several distinct GO clusters, supporting the complex transcriptional changes in alveolar macrophages from smokers. However, cells from COPD patients were seen to be mostly associated with decrease of inflammatory response and regulation of immune response, and loss of antigen processing. This observation is consistent with the depletion in the IFNγ-linked module (**Figure 4.5.4**).

Taken together, by overlaying the WGCNA-determined macrophage activation programs to human *in vivo* tissue macrophages using GSEA, hitherto unexplored biological mechanisms were discovered in alveolar macrophages from smokers and COPD patients.

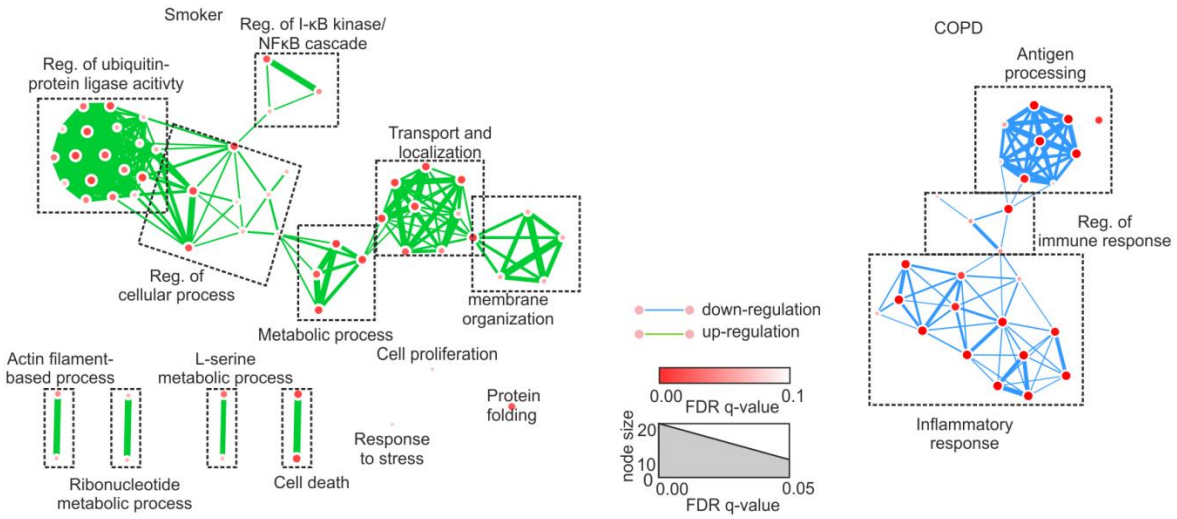

Figure 4.5.4 GO networks for smoker- and COPD-specific modules. Figure was adapted from (Xue et al., 2014). Network visualization of GO enrichment analysis on positively enriched modules (p-value < 0.01) for smokers (modules 41, 6, 21, 16, 39, 5, 19 and 42) and negatively enriched modules (p-value < 0.01) for COPD patients (modules 8, 29) using BiNGO and EnrichmentMap. Node size and node color darkness correspond with the enrichment FDR adjusted p-value (q-value) of the GO term. Green edges stand for up-regulation of the involved genes while blue edges represent down-regulation. Edge thickness shows overlap of genes between neighbor nodes.

4.6 Common activation regulators by network inference with a single method

While the paradigm of macrophage polarization (M1 versus M2) was clearly extended to a multi-dimensional model, this large cohort of data consisting of a variety of stimulation conditions also enabled the identification of common denominators of macrophage activation. To identify these common macrophage activators, the reverse engineering method ARACNe was employed (Margolin et al., 2006). The 9,498 genes present in at least one stimulation condition defined above as informative genes were again used to generate a so-called all-versus-all network (Bonferroni corrected p-value < 10^{-7}) by predicting interactions based on mutual information between each gene pair computed from the expression profiles. As a consequence, 66,744 interactions between 9,073 network genes were identified by ARACNe. This results in an average degree of connectivity of 14.7, which means one gene is involved in 15 transcriptional interactions on average. The parameter settings for reverse engineering calculations and the corresponding network statistics were summarized in **Table 4.6** below:

Table 4.6 Parameters and result summary for reverse engineered networks

Setting No.		i	ii	iii
Condition		all conditions		
No. of samples		299		
Background (mean log2 expression)		6.747431711		
No. of probes		9,498		
Algorithm		ARACNe		TINGe
Parameters	**p-value**	1.00E-07	1.00E-07	1.00E-07
	p-value adjustment	Bonferroni	no	no
	DPI tolerance	0.1	0.1	0.1
Nodes		9,073	9,475	9,485
Edges		66,744	76,075	79,646
Avg. Degree		14.713	16.058	16.794
Regression of degree-node distribution (R-squared)		0.767	0.685	0.710
10 % largest hubs	**Degree cutoff**	30	31	39
	No. of hubs	869	882	909
	No. of interactions	30,431	31,759	35,705
	No. of combined genes (percentage)	0.876777	0.89066	0.884871

To verify the ARACNe network, another reverse engineering approach TINGe (Tool for Inferring Network of Genes) on the basis of information theory was also applied (Aluru et al., 2013). The TINGe network revealed high similarity in the network size and topology as well as the rank of hubs based on degrees of connectivity as determined by ARACNe (**Figure 4.6.1**). Due to the size and complexity of the entire network, instead of displaying it completely, the network statistical properties (degree-node distribution) were summarized in **Figure 4.6.2**.

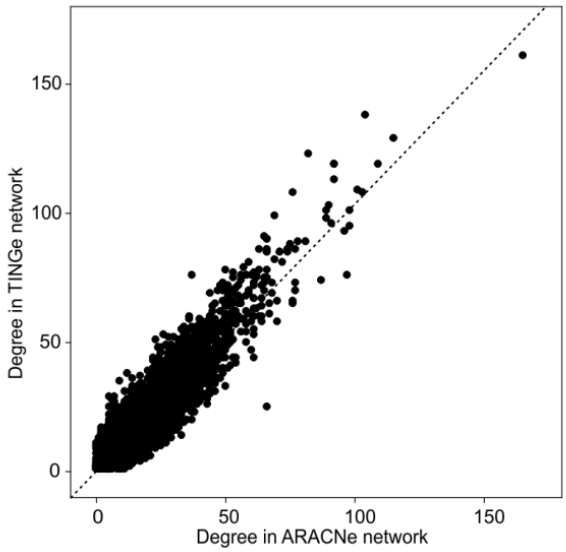

Figure 4.6.1 Comparison of topology of two networks generated from two algorithms: ARACNe and TINGe. Figure was adapted from (Xue et al., 2014). Shown is the degree of connectivity of 9,485 genes within the networks. Parameters used for network generation are identical (Setting No. ii and iii as described in **Table 4.6**).

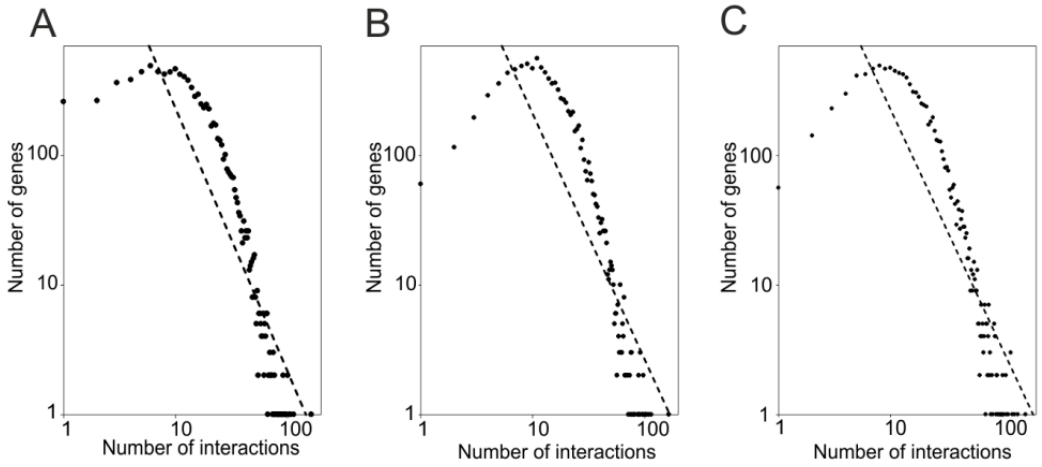

Figure 4.6.2 Degree node distribution of the three major networks generated from different tools or settings. Figure was adapted from (Xue et al., 2014). **A** ARACNe cutoff p-value with Bonferroni correction, **B** ARACNe without Bonferroni correction, and **C** TINGe without p-value adjustment. In each plot, the number of genes with the same number of interactions (from 1 to 164) fits to a power law (dash line) in logarithmic range (R^2 = 0.767, 0.685 and 0.710 for **A**, **B** and **C**, respectively). This indicates that they are scale-free networks.

The top 10% of hub genes (n = 869), which are the most interconnected genes inferred in the network, collectively participated in 30,431 interactions. They were visualized in **Figure 4.6.3**. In the 10 most highly interconnected genes, FABP5 has recently been implicated in lipid

metabolism and inflammation crosstalk (Furuhashi et al., 2011), and TNFAIP6 is a negative feedback regulator of myeloid cell activation (Choi et al., 2011). The next step was to search for current publications on the other most highly interconnected genes using pubatlas.org database. According to the result (**Appendix B Table 1**), little is known about the role of these important network genes in macrophages as well as other immune cells indicating that reverse engineering of large gene expression data from diverse signal inputs reveals unknown aspects of macrophage activation.

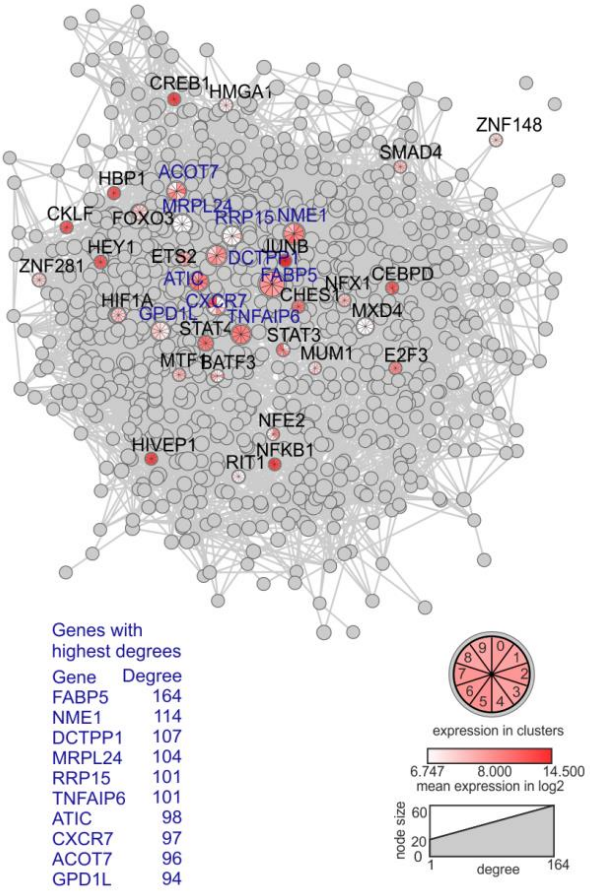

Figure 4.6.3 Visualization of the 10% largest hub genes of the ARACNe predicted macrophage regulatory network. Figure was adapted from (Xue et al., 2014). For the top 10 genes (highest degree of connectivity, blue) and TFs, mean expression values (log2, derived from the 10 clusters in **Figure 4.1.6**) are highlighted in red colors. Node size reflects degree of connectivity.

Next, GO enrichment analysis of the top 10% of hub genes was performed to further understand the biological functions of these important genes (**Figure 4.6.4**). The significantly enriched GO-terms (FDR adjusted p-value < 0.05) were visualized as a network, which was subdivided into 5 major clusters. The biggest cluster was associated with immune response

processes (especially terms associated with 'regulation of activation'). However, other clusters that were related to metabolic and catabolic processes, cell death, biosynthetic processes of small molecules, also constitute major aspects of macrophage activation.

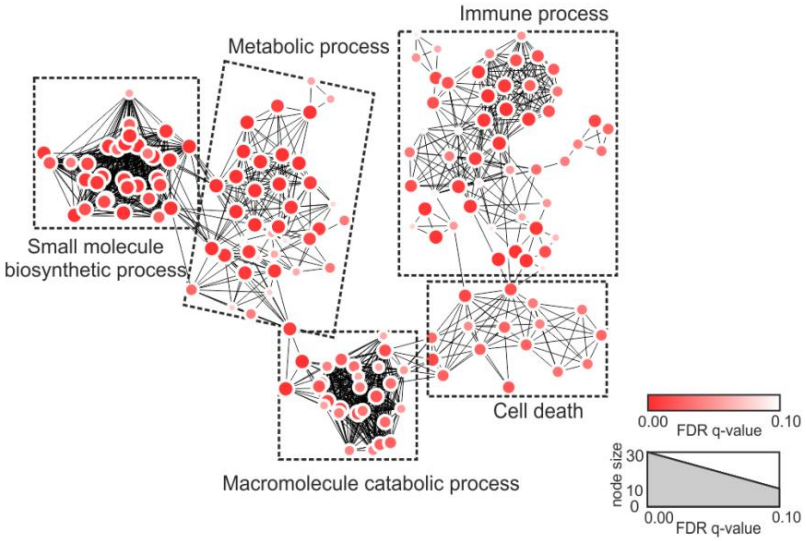

Figure 4.6.4 GO network of 10% largest hubs. Figure was adapted from (Xue et al., 2014). Network visualization of GO enrichment analysis on 10% largest hubs shown in **Figure 4.6.3**. Red nodes represent enriched GO-terms, node size represent FDR-adjusted enrichment p-value (q-value). Edge thickness represents overlap of genes between neighbor nodes.

To identify TFs involved in the common macrophage regulatory network, the top 10% hub genes were intersected with the TFCat database (Fulton et al., 2009) and in the end derived 27 TFs (**Figure 4.6.3**). As the most highly expressed TFs are supposed to be the most relevant players in macrophage activation, and thus the 27 TFs were ranked by their average expression in 29 conditions and a network of the top 5 TFs (JUNB, NFKB1, HIVEP1, CREB1, and HBP1) with their first neighbors within the top 10% hubs (**Figure 4.6.5**). It has been established that these TFs are involved in macrophage activation: JUNB as part of the AP1 complex, NFKB1 (global activator), and CREB1 (inducing survival signals) (Wen et al., 2010), except for the zinc finger protein HIVEP1, whose role is unknown. HBP1 has an impact on differentiation of malignant myeloid cells and the regulation of other important TFs including PU.1, RUNX1, JUNB, or CEBP (Yao et al., 2005). To investigate the potential binding sites of the 27 TFs, position-weight matrices in Genomatix was used to find out the enrichment of gene loci of the top 10% hub genes (**Appendix B Table 2**). 26 out of 27 of these TFs showed significantly enriched binding prediction (p-value < 0.05).

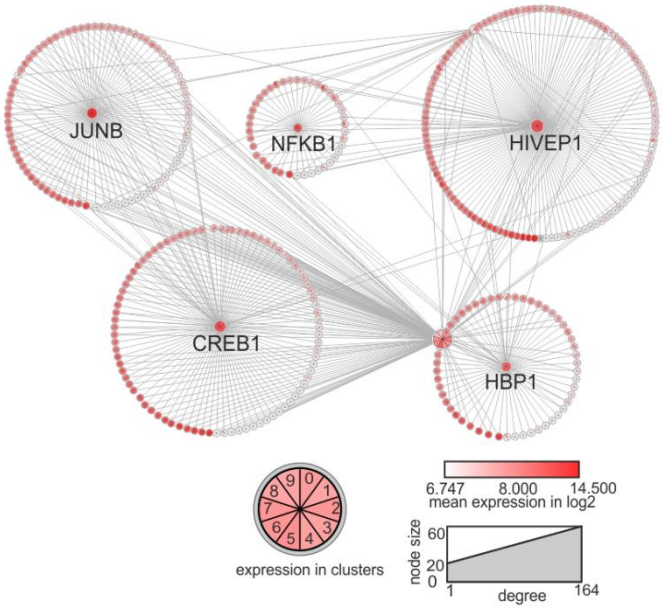

Figure 4.6.5 Sub-networks of the 5 most highly expressed TFs from 10% largest hubs. Figure was adapted from (Xue et al., 2014). First neighbors are surrounding corresponding TFs. Each gene is multi-colored according to its log2-transformed mean expression in 10 clusters from **Figure 4.1.6**.

As a complementary approach, ToppGene (Chen et al., 2009) and Endeavour (Tranchevent et al., 2008), two gene prioritization tools on the basis of machine learning, were applied as well to rank the potential association and closeness of the top 10% hub genes with macrophage cellular programs. For this purpose, the macrophage lineage TFs RUNX1 and SPI1 (PU.1) were used as training set. As a result, the top 11 ranked genes were TFs and in total, 20 of the 30 top ranked genes are associated with transcriptional regulation (**Figure 4.6.6**). Besides NFKB1, JUNB and CREB1, additional TFs already associated with macrophage activation (STAT3) as well as other TFs not yet linked to macrophage activation (e.g. HMGA1, NFE2, ZNF148, SMARCA2, DDX21, MNDA, and TBLX1) turned out to be important. Several macrophage differentiation markers such as CSF1 (M-CSF) and MMP9 were also strongly linked to macrophage activation in this analysis.

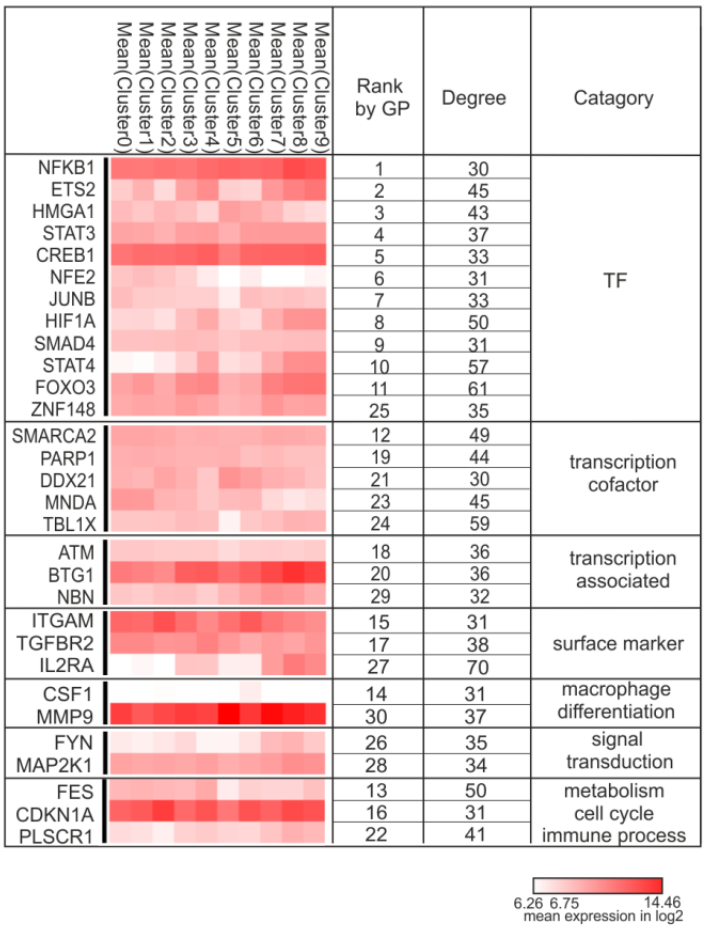

Figure 4.6.6 Top 30 putative candidates after gene prioritization of 10% largest hubs. Figure was adapted from (Xue et al., 2014). Mean log2-transformed expression from each cluster is displayed as a heatmap. The categorization according to cellular functions.

Overall, this gene regulatory network analysis not only confirmed the involvement of known transcriptional regulators such as NFKB1, but also uncovered unexpected and hitherto unexplored candidate regulators, and identified five distinct GO clusters of biological processes as part of the macrophage activation programs.

Taken together, the M1 versus M2 dichotomous system has been extended into a "multi-dimensional model" for human macrophage activation by transcriptome-based network analysis. On the one hand, macrophage activation undergoes stimulus-specific transcriptional regulation, which can be also observed in human tissue macrophages. On the other hand, common regulatory denominators such as NFKB1 play a central role in macrophage activation irrespective of input signals.

5 Refinement of common regulatory networks by integrating multiple inference methods

In the following paragraphs, I illustrate the power of integrating several different approaches of network technologies to derive at a consensus network that outperforms single approaches. This part is prepared as a manuscript for submission.

5.1 The community-based network is a hierarchical scale-free network

To refine the common macrophage activation network, the same transcriptional gene expression profiling data as described in **Chapter 4** (9,498 informative genes from a set of 299 samples, representing perturbation of macrophage phenotypes, ranging from inactivated subpopulation to a variety of stimulated macrophages *in vitro*) were utilized as input for network calculations. Instead of using only one network inference algorithm, community-based methods were applied by integrating complementary network inference approaches. Gene regulatory network prediction was performed by applying five network inference methods: ARACNe, TINGe, CLR, GENIE3 and BasicCorrelation, most of which are proposed to be complementary on performance (Marbach et al., 2012). The prediction settings and network statistics are shown in **Table 5.1**.

Table 5.1 Network statistics on individual and community-based predictions

Inference method	Parameter settings and cutoffs	No. of nodes	No. of edges	Average degree
ARACNe	Bonferroni corrected p-value 1e-7, DPI tolerance 0.1	9,073	66,744	14.713
TINGe	Unadjusted p-value 1e-7, DPI tolerance 0.1	9,486	79,646	16.792
CLR	100,000 edges	7,535	99,999	13.271
GENIE3	100,000 edges	8,372	99,999	19.035
BasicCorrelation	Pearson correlation, 100,000 edges	7,106	99,025	25.744
Community-based (consensus network)	Top 10,000 edges sorted by average rank	4,053	8,504	4.196

Parameter settings and cutoffs used for five individual approaches and community-based methods are shown. The number of nodes, edges and average degree of connectivity of the resulting networks were computed by Cytoscape.

Subsequently, five individual predictions were integrated into one consensus network. The consensus network consists of 4,053 genes and 8,504 edges between them (**Figure 5.1.1**), which resulted in an average degree of connectivity of 4.196, meaning one gene is involved in 4 transcriptional interactions on average.

To further study the topology of the network, the degree-node distribution was visualized as a summary of the network statistical properties (**Figure 5.1.2**). The power law regression in the relationship between the number of nodes (number of genes ranging from 1463 to 1) and their degree (number of interactions ranging from 1 to 76) suggests a scale-free network structure, i.e. the network is unevenly populated with highly connected nodes or hubs and less dense nodes. The R-squared value is 0.942, which is close to 1, indicating high correlation and a strong linear relationship between two variables.

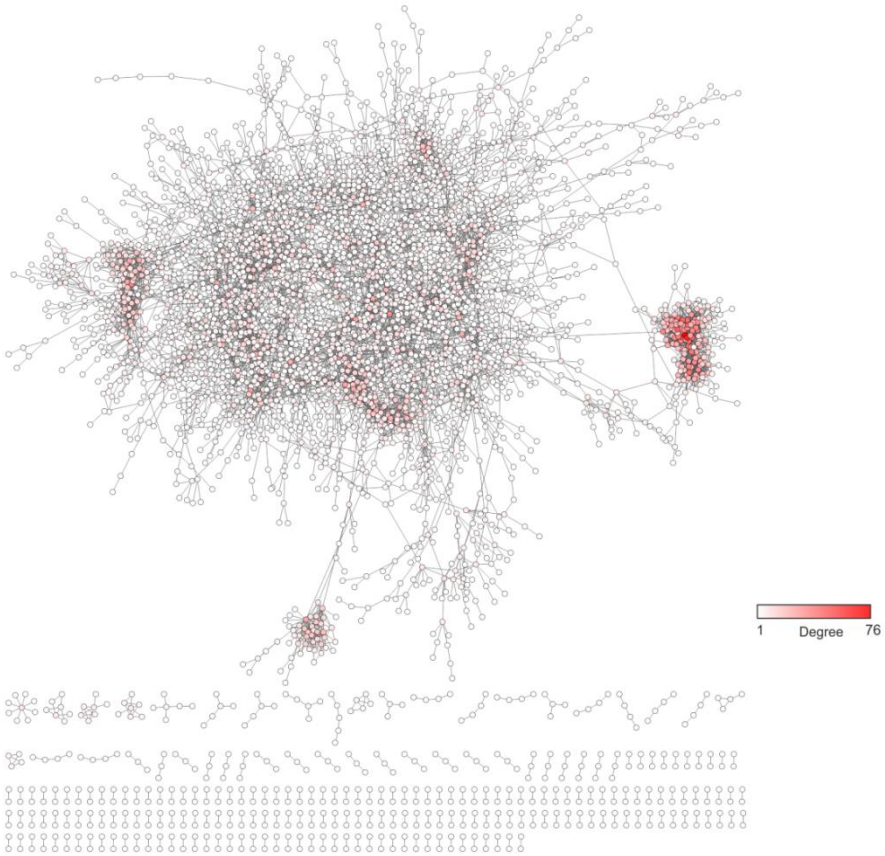

Figure 5.1.1 Complete consensus network and degree-node distribution. Visualization of consensus macrophage regulatory network generated from gene expression profiles (n = 299 transcriptomes) combining five inference methods. Node color and size reflect degree of connectivity.

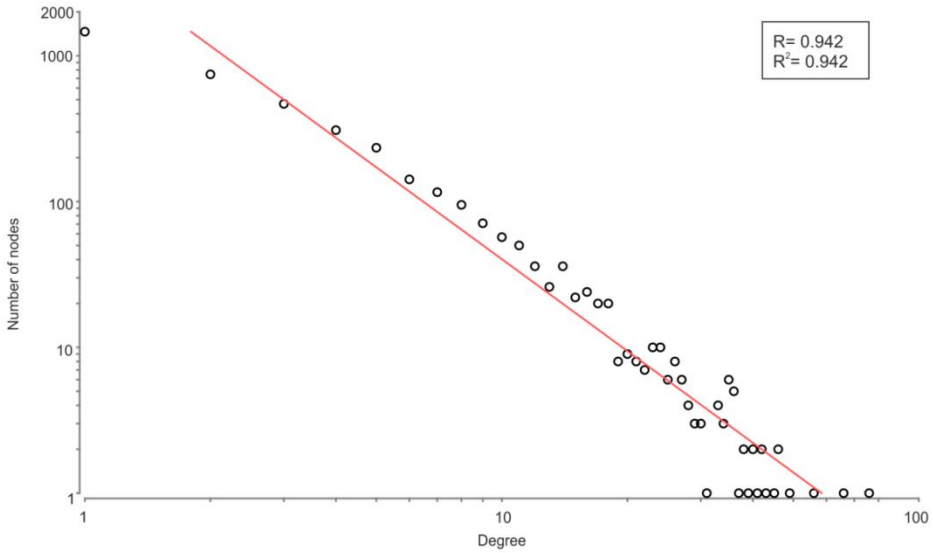

Figure 5.1.2 Degree node distribution of the consensus network. The number of genes with the same number of interactions (from 1 to 76) fits to a power law (red line) in logarithmic range (R = 0.942, R^2 = 0.942). This indicates that they are scale-free networks.

5.2 Transcription factors are the most important network components

In order to select largest central hubs, a threshold to a degree of 10 was set to obtain the 10% top hub genes, i.e., 409 most interconnected genes in the inferred network; they collectively participate in 2,342 interactions, which constitute 27.5% of all edges of the complete network (**Figure 5.2.1**). Interestingly, the 10 most central hubs with highest degrees are CREB1, BLZF1, C9orf80, GJC1, ZNF430, ANKRD30B, NLRP8, SDHALP1, DMC1 and ZMAT3, two of which are transcriptional regulators. Among the top 10% largest hubs, 10.76% are transcription factors (TFs), which is much higher than observed in the ARACNe-inferred network from **Chapter 4.6** (Xue et al., 2014).

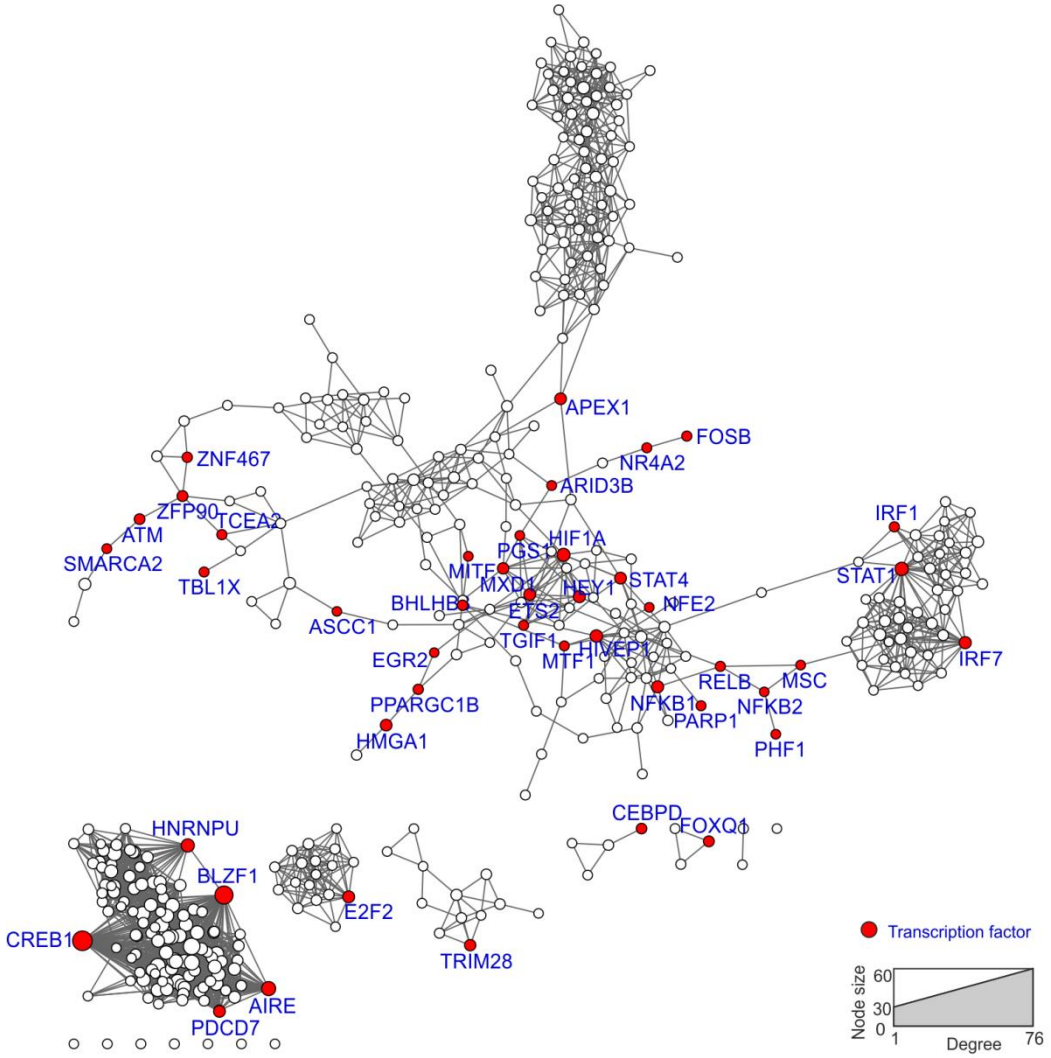

Figure 5.2.1 Top 10% largest hubs of the consensus network. Visualization of the top 10% largest hubs in the consensus macrophage regulatory network. Node size reflects degree of connectivity. The 44 transcription factors are marked in red.

Among the most prominent TFs associated with macrophage activation independent of input signal were CREB1, BLZF1 and AIRE. One TF CREB1 and its first neighbors in the top 10% hubs were extracted and the fold-change values of these genes when comparing IFNγ, IL4 and TPP stimulated macrophages with baselines were visualized by node color (**Figure 5.2.2**). Apparently, there are 73 potential targets for CREB1 and most of the targets were strongly induced by TPP stimulation. Similar visualization with less strong up-regulation could be seen for IFNγ-stimulated cells while only very few of the genes increased their expression in the IL4 condition.

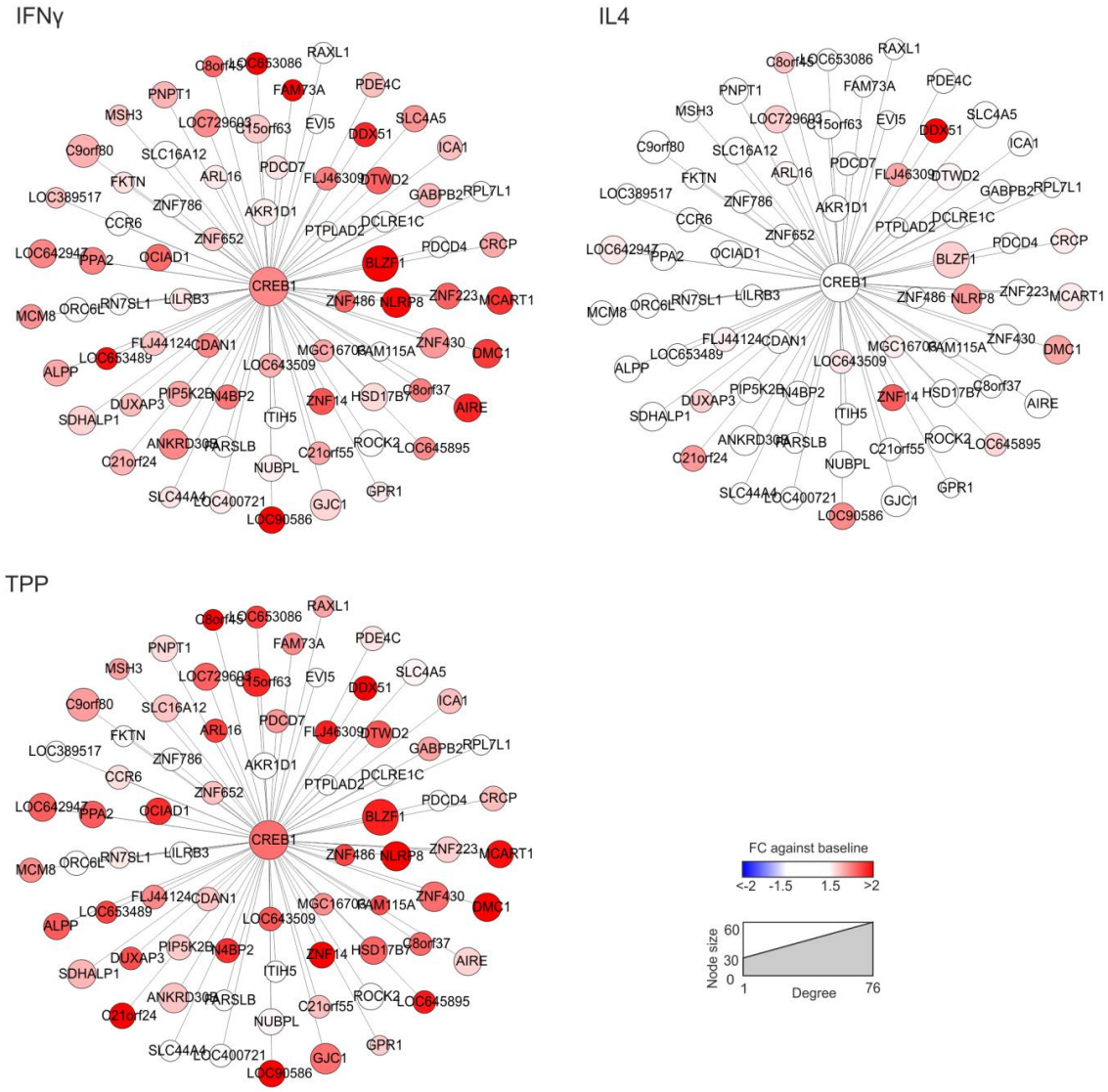

Figure 5.2.2 Transcription factor CREB1 and its first neighbors (potential target genes) within the network. Node color represents FC value of IFNγ-, IL4- or TPP-stimulated macrophages versus baseline while the node size stands for the respective p-value.

5.3 Distinct network clusters reflect very specific cellular functions

The top 10% sub-network was clustered into a certain number of network modules based on topology assigned by the MCL clustering algorithm, and the specific biological functions of these gene clusters were determined by GO terms using BiNGO (**Figure 5.3**). As a result, 18 network modules with at least 5 genes involved were identified. But only larger modules with more than 10 genes showed functional enrichment. The largest module contained 76 genes that were found to be enriched for GO-terms like “Nucleic acid metabolic process”. The second largest module has 44 genes and most of them were associated with innate immune response, type I interferon signaling pathway, etc. Most surprisingly, there was one cluster of genes to be associated with protein targeting to membranes, which was not revealed in such clarity by individual methods.

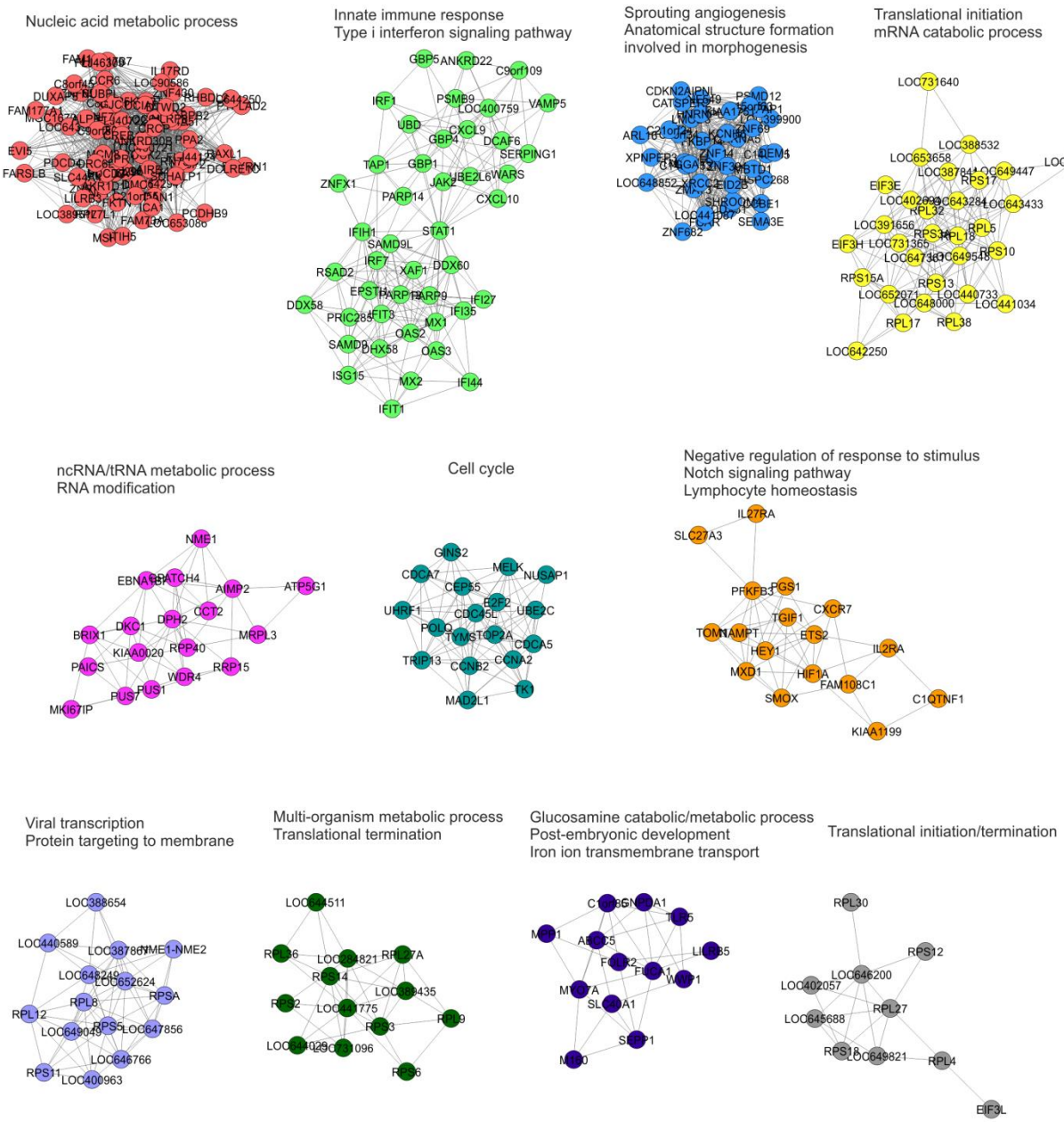

Figure 5.3 Function-specific network modules. 11 identified network modules with at least 10 genes. The enriched GO terms were marked for bigger modules.

Taken together, reverse network engineering of large transcriptional data by integrating several individual methods revealed a TF-centered regulatory circuit with unexpected complex functional components during macrophage activation.

6 Transcriptome-based network analysis refines pathway models for host defense against bacterial and fungal pathogens

In this chapter, I present an example, how not only the methods, but also the information within the macrophage data can be applied to a dataset derived from a more complex mixture of cells, namely peripheral blood mononuclear cells (PBMC). PBMC are regularly studied in clinical settings addressing their reactivity towards exogenous stimuli. During the last 10 years numerous studies on reactivity of PBMC to pathogens have been studied, one of the largest being reported by one of our collaborators Prof. Mihai G. Netea from the Radboud University in Nijmegen, The Netherlands. I used this dataset to illustrate, how activation of blood-derived monocytes, precursors of macrophages can be studied in the context of a complex mixture of cells. This chapter is also prepared for submission as a manuscript.

6.1 The transcriptomic relationships between human PBMC stimulated with microbial pathogens

Human PBMC of 46 health donors were isolated from peripheral blood and stimulated with four types of microbial stimuli. Their transcriptomes at 4 and 24 hours were assessed by microarray experiments and compared globally. Using co-regulation analysis to assess overall sample-to-sample relationships, cells from earlier time points were clearly distinguishable from cells stimulated after a longer time (**Figure 6.1.1**), which was confirmed with other approaches such as principal component analysis and hierarchical clustering (data not shown). Cells stimulated with LPS, *B. burgdorferi* (Bbu) and *M. tuberculosis* (Mtu) were clustered next to the unstimulated cells while *C. albicans* (Cal) treated PBMC are localized farther away from their steady state for both time points. At 24 hours, the Cal treated cells even abrogate the initial axis and extend into a third axis (**Figure 6.1.1**).

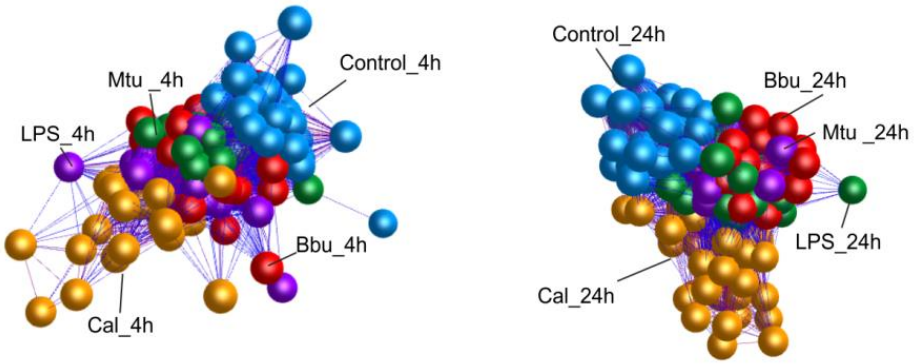

Figure 6.1.1 Sample-sample co-regulation network of 294 transcriptomes from PBMC. The Pearson correlation threshold of 0.97 was used. Each node stands for one transcriptome. Nodes are colored with conditions and time points.

6.2 The cellular composition of human PBMC stimulated with microbial pathogens

Very recently, cell-type identification by estimating relative subsets of RNA transcripts (CIBERSORT), a computational approach that accurately resolves relative fractions of diverse cell subsets in gene expression profiles (GEPs) from complex tissues, was introduced (Newman et al., 2015). CIBERSORT was applied using the previously published gene signatures for cells within the PBMC compartment (Newman et al., 2015) to the 294 GEPs derived from PBMC stimulated with Cal, Bbu, Mtu, and LPS for 4 and 24 hours prior GEP (**Figure 6.2**). The complete transcriptomes were first filtered to retain 12,029 unique present genes by testing if the probes are expressed in at least one condition using a calculated background value of 6.87. At the 4 hour time point, PBMC stimulated with Cal showed the most profound changes in gene signature composition with an enlargement of the gene signatures for activated memory CD4 T cells, NK cells and even the appearance of a $M^{IFN\gamma+sLPS}$ macrophage signature. These findings are in line with previous observations demonstrating IFNγ and sLPS to be a central player in the defense against Cal (Smeekens et al., 2013). However, the CIBERSORT-based model further suggested that the global IFNγ and sLPS signature is due to at least three independent cellular compartments, namely CD4+ T cells, NK cells and monocyte-derived macrophages ($M^{IFN\gamma+sLPS}$). Bbu, Mtu and LPS-stimulated PBMC did not show similarly prominent differences at the 4h time point. At the 24h time point, however, a relative increase of the baseline macrophage (M^{b}) and the activated DC signature as well as the regulatory T cell (Treg) signature in Bbu-, Mtu-, and LPS-stimulated PBMC was observed. This was in contrast to Cal-stimulated PBMC that were characterized by a further increase of $M^{IFN\gamma+sLPS}$, activated NK cells and CD4+ T cells and a complete lack of the Treg

signature. Collectively, CIBERSORT-based modeling of *in vitro* stimulated PBMC revealed the myeloid compartment to be mainly involved in response to bacterial stimuli within the first 24 hours, while at least three cellular compartments were involved in the response against Cal.

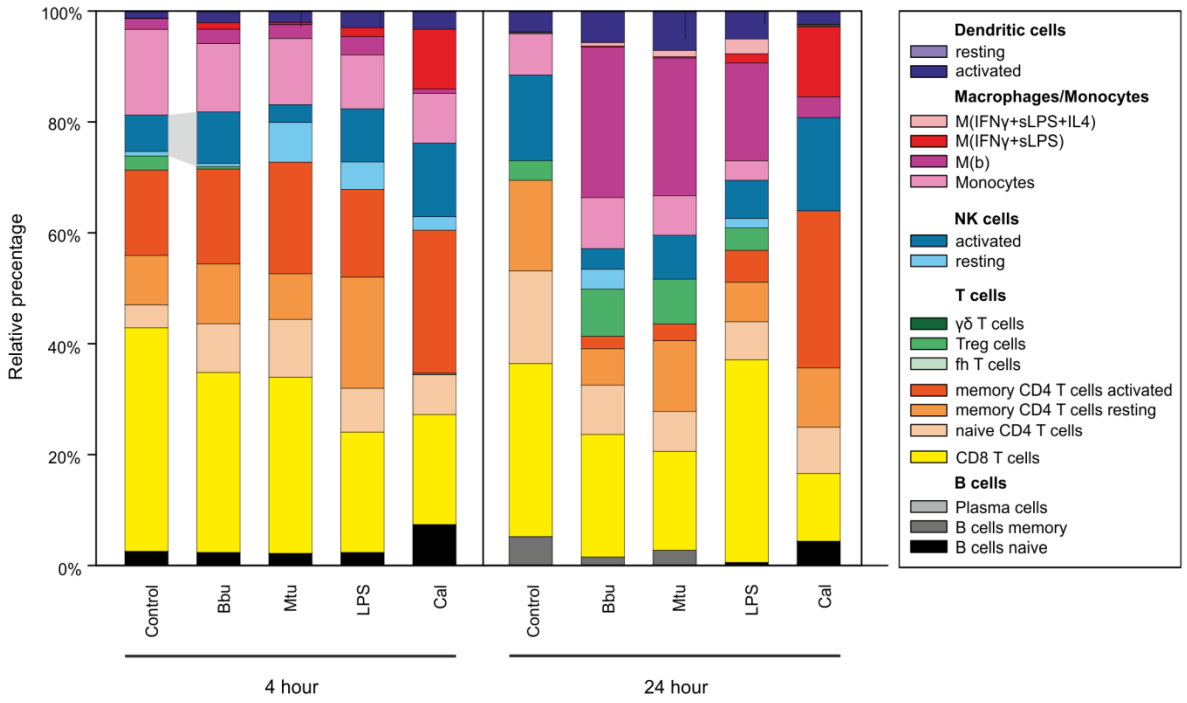

Figure 6.2 Relative fractions of human hematopoietic cell types for the 10 different conditions within the dataset. The estimated abundances (relative fractions) of member hematopoietic cell types within the dataset are visualized as a stacked bar plot. As human leukocyte signature matrix, 18 cell phenotypes including naïve and memory B cells, plasma cells, 7 T cell subsets, resting and activated NK cells and myeloid compartments, were used. M(b): inactivated macrophages; M(IFNγ+sLPS): macrophages activated by IFNγ and sLPS ; M(IFNγ+sLPS+IL4): macrophages activated by IFNγ, sLPS and IL4; Treg cells: regulatory T cells; fh T cells: follicular helper T cells.

6.3 Common transcriptional programs and molecular pathways for infections by Bbu, LPS, Mtu and Cal

To identify common response genes after four types of microbial stimulation, an ANOVA model was applied for each stimulus versus the corresponding control to obtain differentially expressed (DE) genes from 12,029 unique present genes as described above. By overlaying the DE genes for four stimuli at two time points, 234 and 325 common DE genes were identified for 4 and 24 hours, respectively. And in them, 86 genes were shared by both time points (**Figure 6.3.1**). The large majority of DE-genes, including pro-inflammatory cytokines

such as CCL2, CCL3, CCL20, CXCL1, CXCL5, IL1A and IL1B, as well as transcription regulators such as NFKBIA and NFKBIZ, were induced by bacterial and fungal pathogen stimulation whereas only four of them, namely VCAN, FGL2, FCER1A and RNASE6, were inhibited by the stimuli (**Figure 6.3.2**).

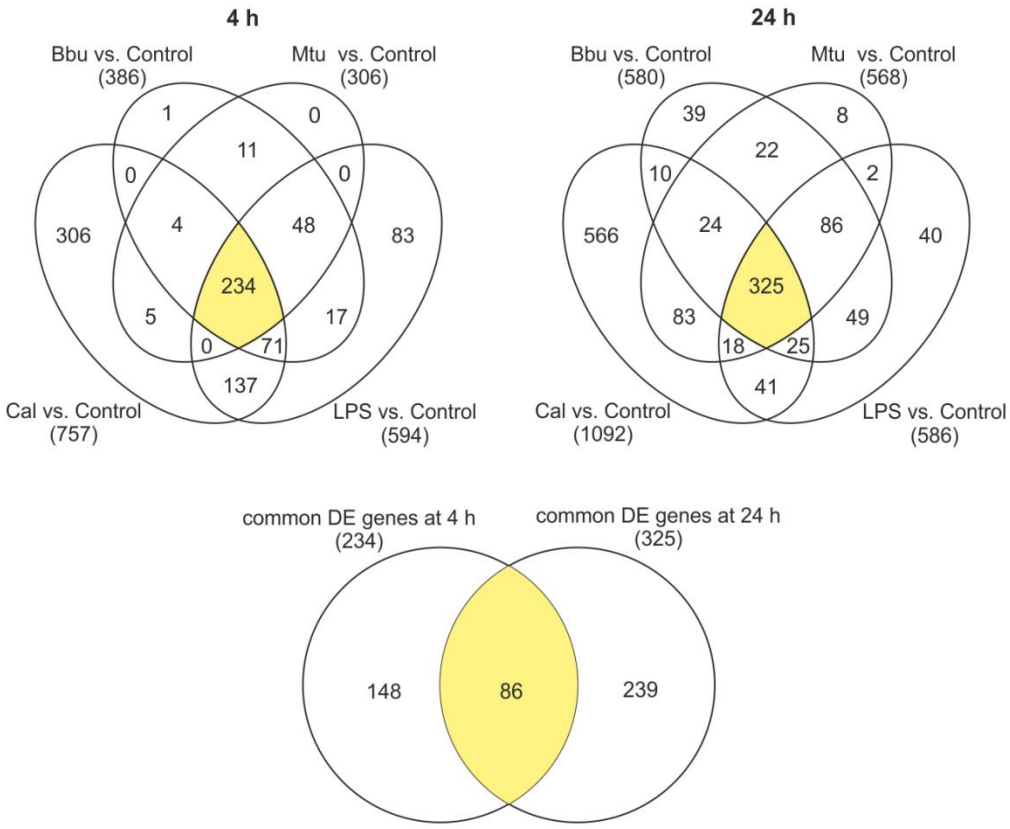

Figure 6.3.1 Venn diagram of identification of 86 common differentially expressed (DE) genes induced by four microorganisms. The DE genes for each comparison were identified with |FC| > 1.5 and FDR adjusted p-value < 0.05 from 12,029 present genes.

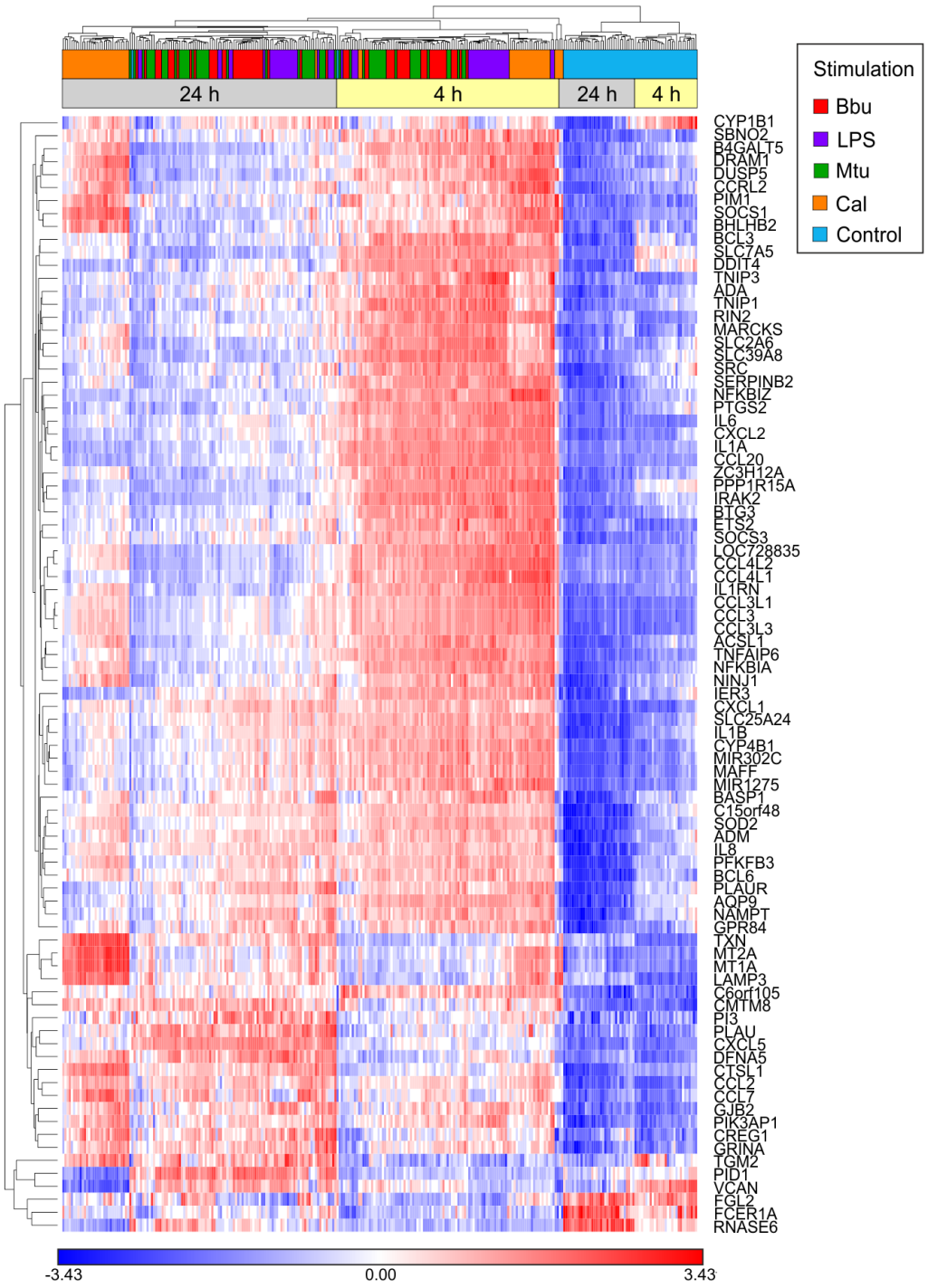

Figure 6.3.2 Hierarchical clustering on 86 common differentially expressed genes induced by four microorganisms. Heatmap displays z-transformed expression values (z-scores) from blue via white to red.

To understand the biological functions of these common response genes, GO enrichment analysis was performed and followed by network visualization of enriched GO-terms using BiNGO and EnrichmentMap (**Figure 6.3.3**). This analysis demonstrated the major functional processes of PBMC reacting to microorganisms are for example inflammatory response, defense response, acute inflammatory response and ion homeostasis. Pathway enrichment analysis confirmed the common effective pathways for host defense against these microorganisms (**Table 6.3**).

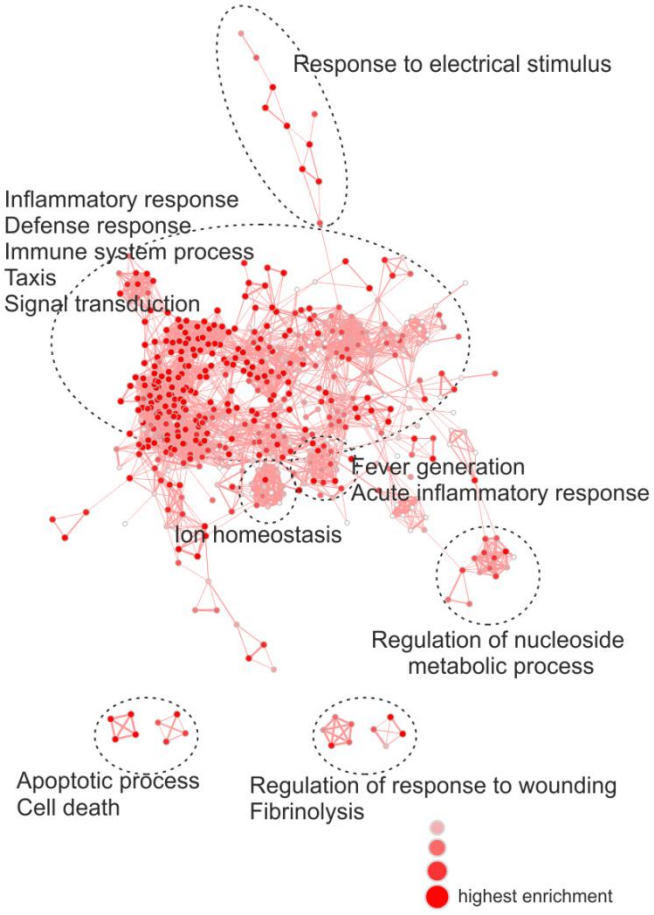

Figure 6.3.3 GO network of 86 common differentially expressed genes induced by four microorganisms. Each node represents one enriched GO term (FDR q-value < 0.05). The similarity cutoff was Jaccard coefficient of 0.45. Node size and node color darkness correspond with the enrichment FDR adjusted p-value (q-value) of the GO term, i.e. the biggest and darkest nodes are the GO terms most significantly enriched with lowest FDR q-value. Edge thickness shows overlap of genes between neighbor nodes.

Table 6.3 Pathway enrichment analysis on 86 common differentially expressed genes

Pathway Name	Enrichment Score	Enrichment p-value	% genes in pathway that are present
Rheumatoid arthritis	26.6907	2.56E-12	13.1868
Salmonella infection	23.875	4.28E-11	12.5
NOD-like receptor signaling pathway	15.0457	2.92E-07	11.8644
Legionellosis	12.4082	4.08E-06	10.7143
Cytosolic DNA-sensing pathway	9.10482	0.00011113	8.06452
Pertussis	8.07669	0.0003107	6.49351
Malaria	7.36131	0.00063536	7.84314
Epithelial cell signaling in Helicobacter pylori infection	6.27754	0.00187801	5.88235
Leishmaniasis	5.9663	0.0025637	5.40541
Prion diseases	5.91636	0.00269498	8.33333
Graft-versus-host disease	5.54302	0.00391469	7.31707
Apoptosis	5.34007	0.00479555	4.54545
Shigellosis	4.43566	0.0118473	4.91803
Adipocytokine signaling pathway	4.06532	0.0171575	4.28571
Hematopoietic cell lineage	3.49676	0.0302953	3.44828
African trypanosomiasis	3.34176	0.0353745	5.26316
Type I diabetes mellitus	3.11603	0.044333	4.65116

Only significantly enriched (p-value < 0.05) KEGG pathways are shown.

Collectively, these data demonstrate that microbial stimulations cause whole-genome transcriptional changes over time and there are common inflammatory pathways playing roles in PBMC activation induced by four types of microorganisms.

6.4 Stimulation-specific gene regulatory networks

To better understand the complexity of overall transcriptional regulation of PBMC after microbial infection, 1,690 genes that are differentially expressed (|FC| > 1.5, FDR adjusted p-value < 0.05) in at least one stimulation condition were identified among 12,029 present genes. With a Pearson correlation cutoff of 0.83, a gene regulatory network with the most co-expressed genes (n = 934) was built. The genes with expression changes more than 1.5-fold for each of the four stimulation conditions and two time points as well as the genes differentially expressed uniquely in the corresponding condition, were highlighted in the respective network (**Figure 6.4.1**).

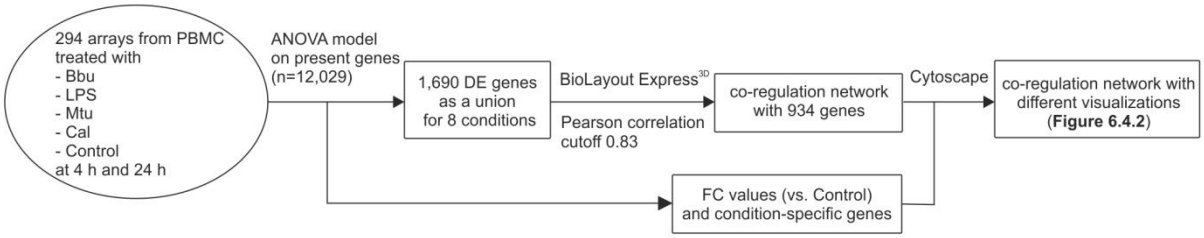

Figure 6.4.1 Schema for co-regulation network analysis of differentially expressed (DE) genes for 8 conditions.

At the 4 hour time point, Bbu and Mtu showed very similar changes with either no or one gene (JUNB) specifically changed after Mtu respectively Bbu stimulation (nodes with black border). The majority of the similarly up-regulated genes (red nodes) were co-localized in one network cluster within the network (**Figure 6.4.2**). When assessing LPS-stimulated PBMC, a second cluster with up-regulated genes was apparent and several genes specifically induced only after LPS treatment (nodes with black border). Moreover, there were a larger number of decreased genes located in two distant network clusters. The most prominent changes were observed in Cal-stimulated PBMC showing the largest number of specifically regulated genes. However, induced or elevated genes were located within the same two network clusters marked in LPS-stimulated cells and this was similarly true for down-regulated genes. At the 24h time point, much more diversity between the stimuli was observed. In Bbu-stimulated PBMC, 28% of the genes regulated in the main network cluster of the 4 hour time point were not elevated anymore, but many additional genes in other regions of the network showed elevation and the number of genes specifically regulated after 24h of stimulation with Bbu also increased. Two distinct clusters of down-regulated genes appeared, one which was similarly down-regulated in all conditions. Overall, the distribution of expression changes of Bbu-stimulated PBMC was most similar to LPS- followed by Mtu-stimulated PBMC. Similar to the 4 hour time point, Cal-stimulated PBMC were characterized by many more genes being differentially and even specifically regulated. Furthermore, there were significant differences in distribution of genes with expression changes throughout the network.

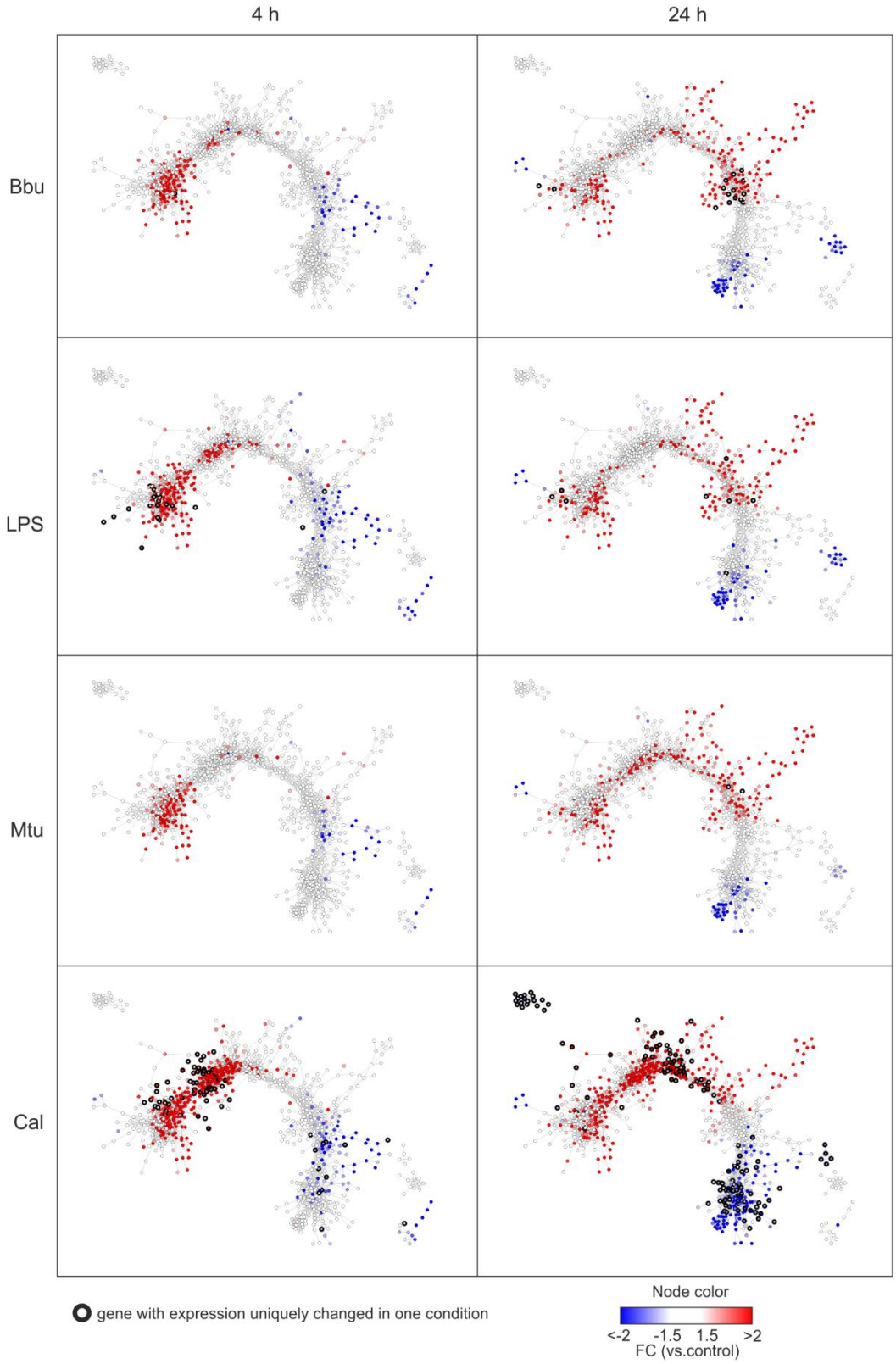

Figure 6.4.2 Network visualization of differentially expressed genes for 8 conditions with condition-specific regulation information. Only those genes with expression changes more than 1.5-fold in each stimulation condition were colored in red (up-regulation) or blue (down-regulation). The genes only showing expression changes in one condition are highlighted with black border.

Next, a co-regulation network was constructed including 572 human transcriptional regulators (TRs), which are present in the gene expression dataset. The most connected TRs were visualized as a network (Pearson correlation cutoff 0.7) and differential expression of TRs for the four stimulation conditions and two time points separately was mapped onto the network (**Figure 6.4.3**). The network was partitioned into two larger and two smaller clusters. Throughout all 8 conditions, 7 commonly induced TFs (NFKBIA, NFKBIZ, ETS2, MAFF, BCL3, BCL6 and BHLHB2; FC > 1.5, FDR adjusted p-value < 0.05) were identified regardless of stimulus and time point (**Figure 6.4.4**) and one of the smaller clusters contained 4 TFs (LYL1, ZNF467, RXRA and CEBPD) that were down-regulated at the 4 hour time point in all conditions. The stimulus-independent regulation suggested that these TRs are involved in common cellular activation programs. Other than common regulators, stimulus-specific TFs within the network were also identified. For example, JUNB, an AP-1 TR, was found to be only induced by Bbu stimulation after 4 hours while this induction could not be observed at 24 hours. STAT1A, GADD45A, ETV3 and ZFAND5 were only differentially expressed (|FC| > 1.5, FDR adjusted p-value < 0.05) 4 hours after LPS treatment while 20 hours later MN1 and PGS1 expression was specifically increased. Due to very similar regulation in Mtu, Bbu and LPS conditions, no unique regulation was observed for Mtu stimulation. In contrast, many more specific up- and down-regulated TRs (|FC| > 1.5, FDR adjusted p-value < 0.05) including NCOA7, IRF4 and RBL2, were identified for Cal-stimulated cells at both time points.

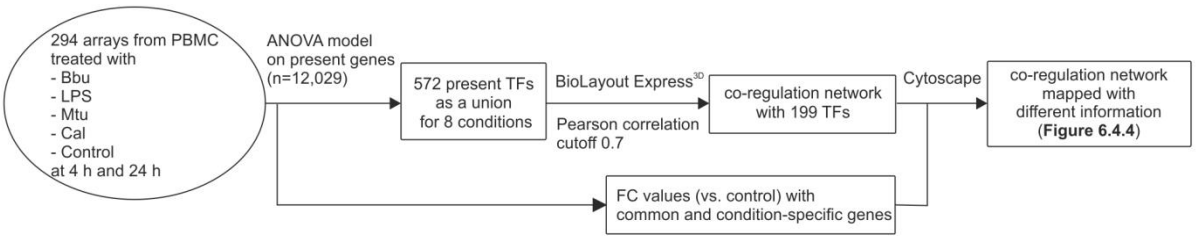

Figure 6.4.3 Schema for co-regulation network analysis of common and condition-specific TFs.

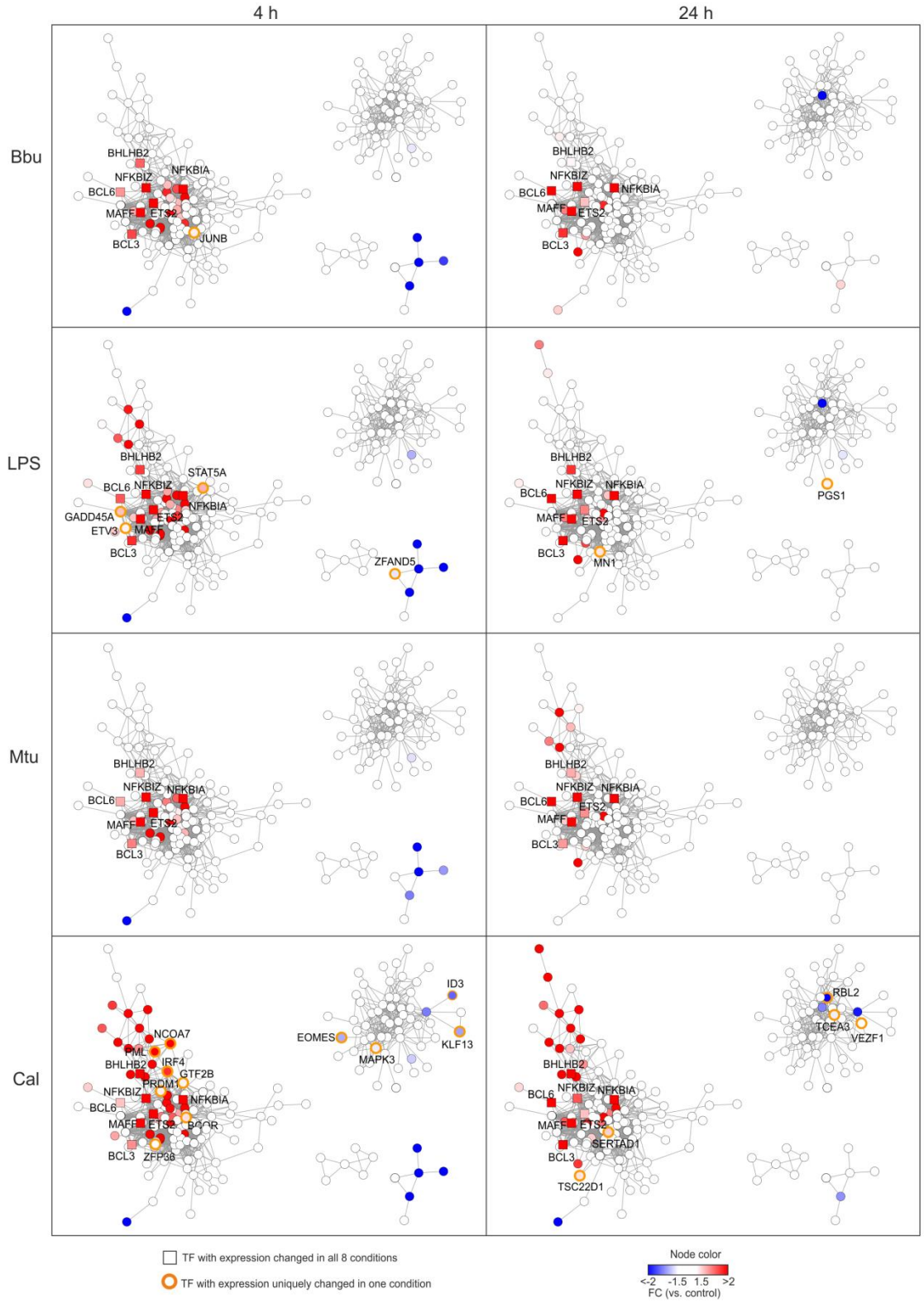

Figure 6.4.4 Network visualization of common and condition-specific TFs. Co-regulation network of TFs that are present in 8 conditions**.** Only those TFs with expression changes more than 1.5-fold in each stimulation condition were colored in red (up-regulation) or blue (down-regulation). Rectangles are DE TFs shared by all 8 conditions. The TFs only showing expression changes in one condition are highlighted with orange circles.

To obtain a global view of the regulation of these TFs, all their FC values for the 8 conditions were summarized within one network by using multiple node colors as shown in **Figure 6.4.5**. Similarity and dissimilarity analysis throughout conditions based on 199 network TFs were performed by Spearman (Rank) correlation matrix and hierarchical clustering on FC values as shown in **Figure 6.4.6** and **Figure 6.4.7**, respectively. Despite dissimilarities between earlier and later time points, the large majority of the up- and down-regulated TRs are separated in two big clusters, left and right, respectively.

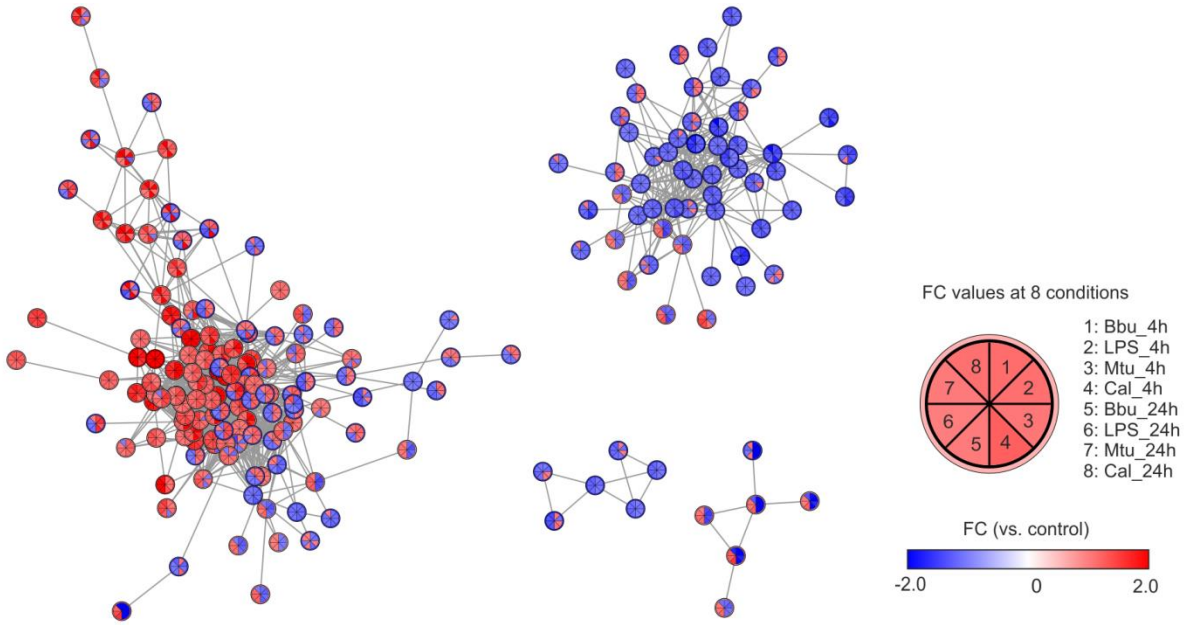

Figure 6.4.5 Network visualization of TFs and their regulation in 8 conditions. Each node is divided into 8 pieces, which display the FC values of 8 stimulation conditions against corresponding controls in blue-to-red color scale.

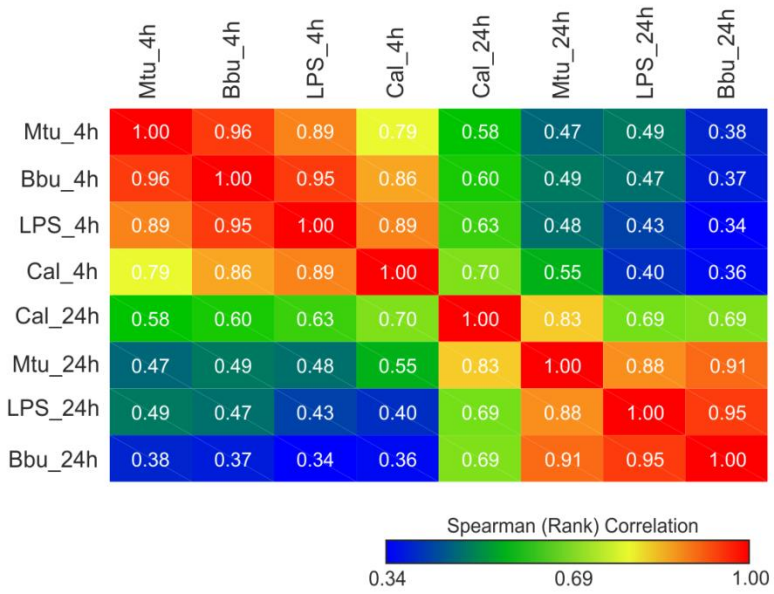

Figure 6.4.6 Spearman correlation coefficient matrix of 8 conditions based on expression changes throughout conditions for network TFs described in Figure 6.4.4-6.4.5. Red colors stand for higher correlation while blue colors represent lower correlation. Yellow and green colors show the medium values.

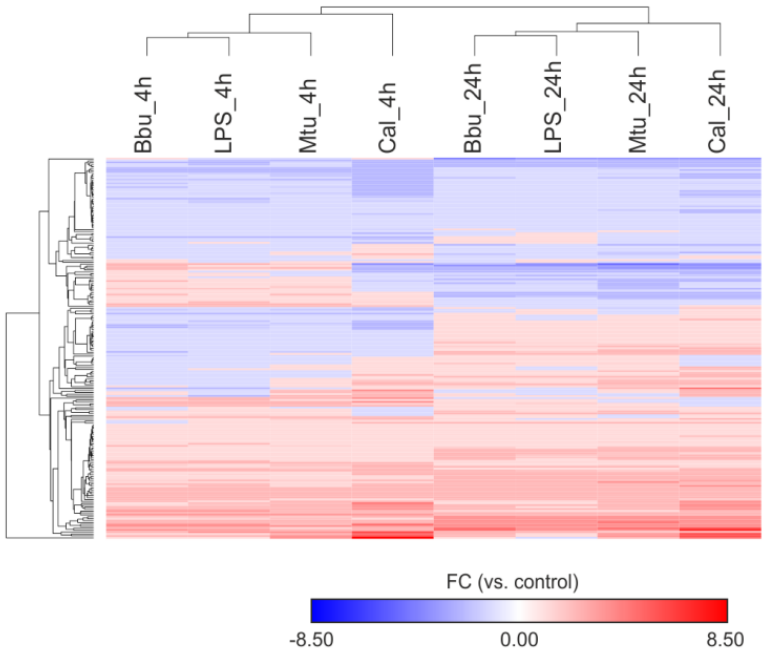

Figure 6.4.7 Hierarchical clustering of 8 conditions based on expression changes throughout conditions for network TFs described in Figure 6.4.4-6.4.5. The heatmap shows z-transformed FC values against corresponding controls in red-to-blue color scale.

Moreover, Cal stimulation showed completely different areas of highly induced signature genes with respect to the bacterial stimuli. In order to address the common and stimulus-specific marker genes, cell surface markers (de Souza et al., 2012), cytokines and chemokines

that are present in the dataset were extracted and the up-regulated marker genes were defined in the ANOVA model (FC > 1.5 and FDR adjusted p-value > 0.05 when comparing to respective controls). As a result, 236 up-regulated surface makers and 55 present cytokines and chemokines (Costantini et al., 2009) were visualized in Circos plots (**Figure 6.4.8-6.4.9**).

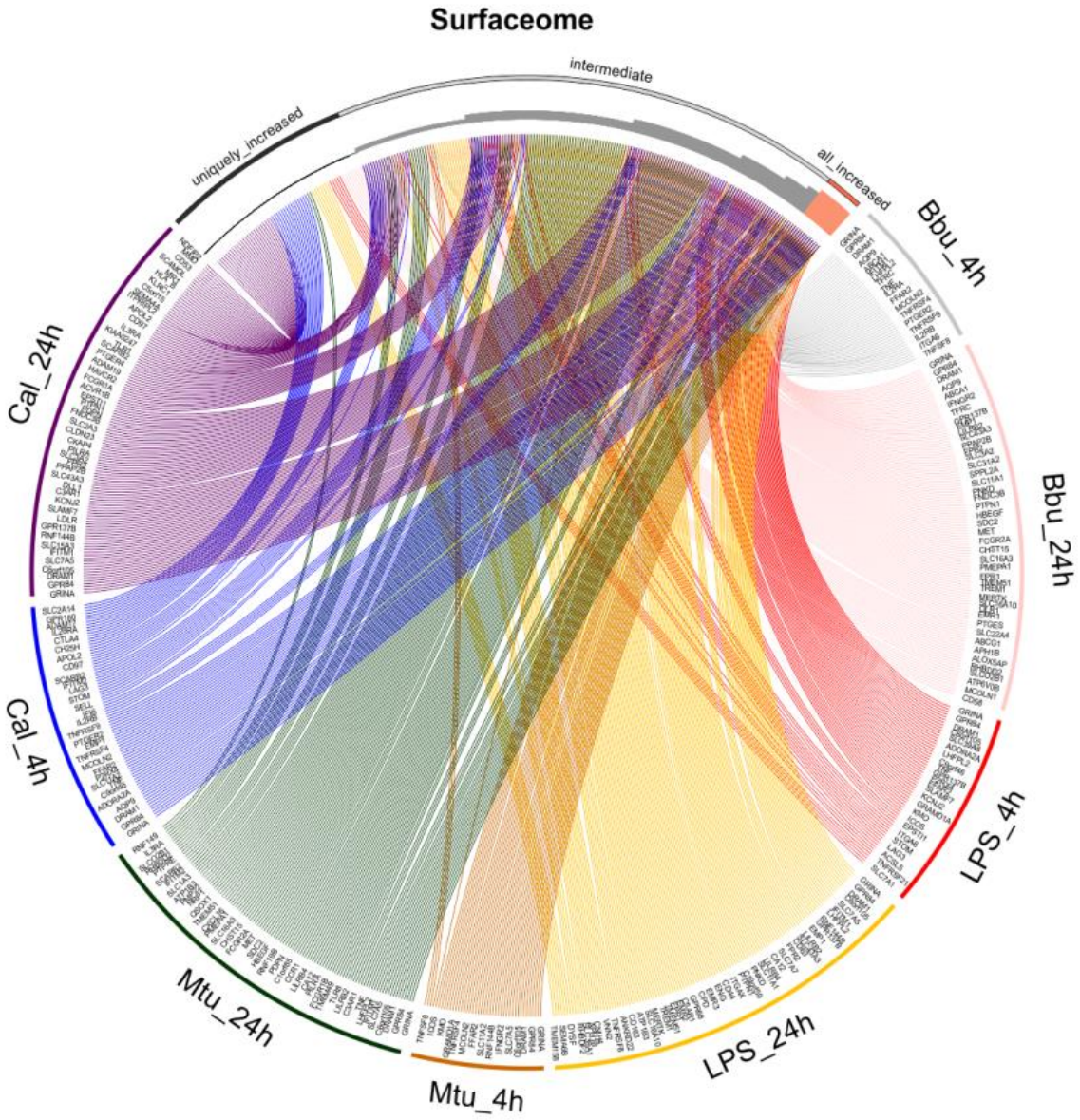

Figure 6.4.8 Circos visualization of shared and condition-specific surface molecules. The histogram on the top shows the number of conditions that show an up-regulation for each surface marker individually for 236 up-regulated surface markers within the dataset. Orange bars represent surface markers that are induced by all 8 conditions while black bars stand for the ones that are only up-regulated in one condition, and grey ones are intermediate between the two. The link from each position of the histogram to one position at any of the 8 conditions stands for one particular surface marker that is increased in the corresponding condition.

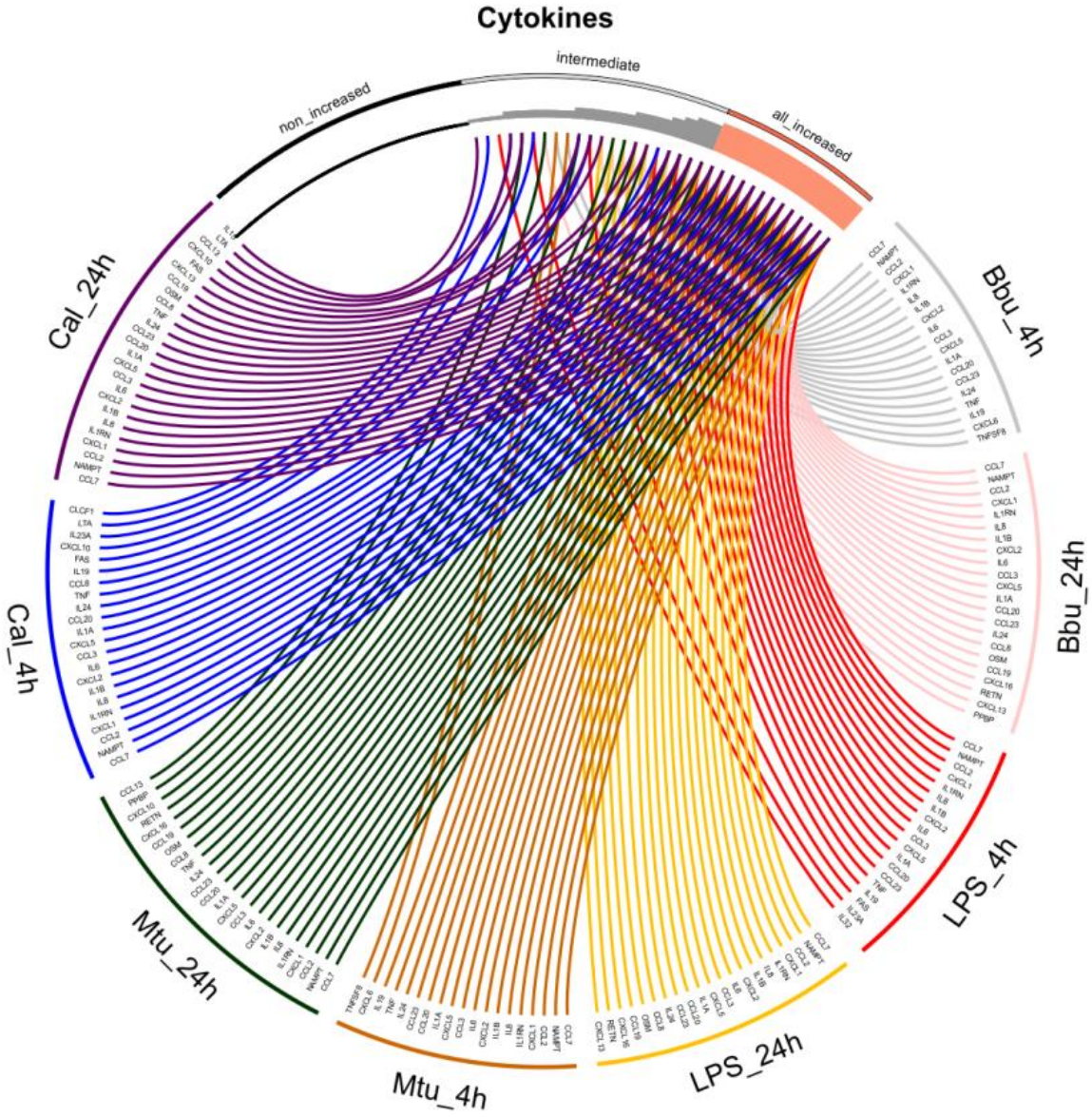

Figure 6.4.9 Circos visualization of shared and condition-specific cytokines. The histogram on the top shows the number of conditions that show an up-regulation for each cytokine individually for 55 present cytokines within the dataset. Orange bars represent ones that are induced by all 8 conditions while black bars stand for the ones that are not up-regulated in any of the 8 conditions, and grey ones are intermediate between the two. The link from each position of the histogram to one position at any of the 8 conditions stands for one particular cytokine that is increased in the corresponding condition.

6.5 Link PBMC data to human macrophage activation signatures

It has been demonstrated in **Chapter 6.2** that the transcriptional signals of different types of macrophages (approximately 10-40%) are a major component of the overall PBMC signature in response to microbial stimulations (**Figure 6.2**). Hence, the previously established information on macrophage activation, namely the 29 human *in vitro* macrophage conditions described in **Chapter 4** (Xue et al., 2014), was linked to the signatures in the PBMC dataset using CIBERSORT. As a result, a signature matrix with 244 genes that discriminate the gene expression of these 29 conditions was generated. On the basis of the macrophage activation signature, estimation of the abundance of the macrophage activation conditions within GEPs derived from PBMC treated with 4 microbial stimuli and RPMI control for 4 and 24 hours was performed, but only the data for 8 stimulated conditions are shown in **Figure 6.5**. As expected, in the Cal conditions at both time points IFNβ and IFNγ signatures were most highly enriched, which confirmed previous findings (Smeekens et al., 2013). At the earlier time point, an sLPS associated activation signature ($M^{sLPS+IC}$) was the largest individual macrophage activation signature within the three bacterial conditions. At the later time point, the upLPS associated signature and signatures associated with TNF (M^{TPP}, $M^{TPP+IFN\beta}$) were elevated. Surprisingly, the HDL signature made up over 10% of the PBMC signature in the Bbu and LPS conditions. Together, these results indicated that although human host defense by PBMC show similar gene response patterns on transcriptional level, linking macrophage activation programs reveals differential gene regulation within the myeloid compartment in the four stimulated PBMC conditions. Collectively, by applying human macrophage activation signatures to admixture population of PBMC, a hitherto unexplored biology in host defense against microorganisms has been uncovered.

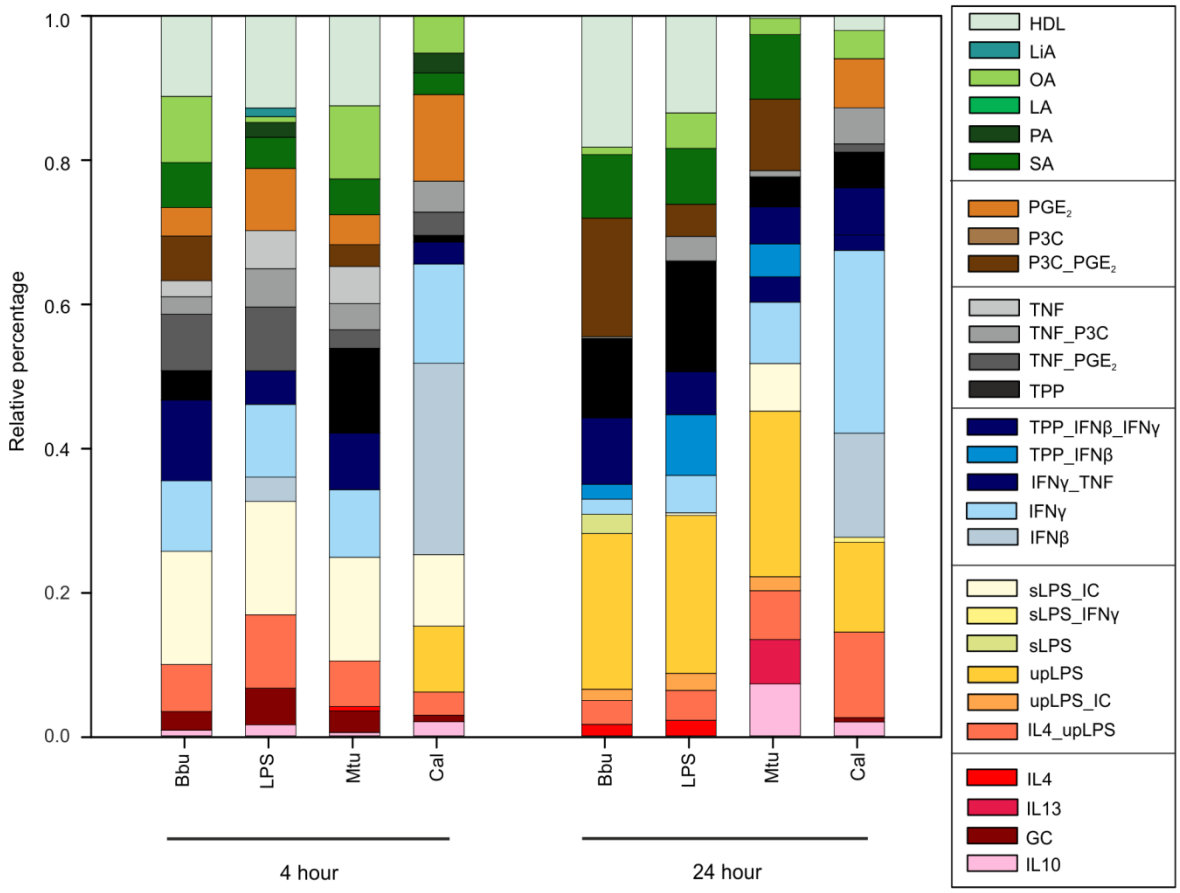

Figure 6.5 Relative fractions of human macrophage activation signature for 8 stimulation conditions. The estimated abundances (relative fractions) of human macrophage activation signatures (28 activation conditions) within the dataset are visualized as a stacked bar plot. As human macrophage activation signature matrix, 184 transcriptomes representing 29 conditions from the human macrophage activation resource data (**Chapter 4**) (Xue et al., 2014) were used for CIBERSORT. GC: glucocorticoid; IC: immune complexes; PGE_2: prostaglandin E2; P3C: Pam3CysSerLys4; TPP: TNF+PGE_2+P3C; PA: palmitic acid; OA: oleic acid; LA: lauric acid; LiA: linoleic acid; SA: stearic acid; sLPS: standard lipopolysaccharide; upLPS: ultrapure lipopolysaccharide; HDL: high density lipoprotein.

6.6 Refined model of regulatory pathways of bacterial and fungal stimulation

Since the myeloid compartment was a major cellular compartment reacting towards microbial stimulation of PBMC, gene regulation was mapped onto the pattern recognition receptors (PRR) pathways known to be involved in the protective effects against these microbial infections in early (4 hours) and late (24 hours) phases. It is important to note that gene expression (as heatmaps) (**Figure 6.6**) and not pathway activation on protein level was mapped onto the pathway maps. Nevertheless, gene expression changes of genes involved in PRR signaling will have a significant impact on the downstream signaling events and therefore can reveal important information about differential usage of known signaling cascades (Arguello et al., 2015). From previous studies, it is known that LPS is exclusively recognized by Toll-like receptor (TLR) 4 (Hoshino et al., 1999). Bbu can be recognized by TLR2 and the NOD2 receptor (Berende et al., 2010). Mtu is mainly recognized by TLR2 and NOD2 receptor involved pathways, in which TLR4 and DC-SIGN also play a minor role (Kleinnijenhuis et al., 2011). As a fungal pathogen, Cal is mainly recognized by C-type lectin receptors (e.g. dectin-1, dectin-2 and mannose receptor) while TLR2 and TLR4 are also involved to some extent (van den Berg et al., 2012). Here, it is now shown at the early time point, that the NFκB and NOD2 signaling pathways were commonly activated by bacterial and fungal pathogens, which led to the expression of pro-inflammatory cytokines (e.g. IL1β, IL6, CXCL2, etc.). However, the overall increase of p105 and Tpl2 resulted in the inhibition of the TF AP-1. More specifically, comparing to bacterial stimuli, Cal stimulation induced the MyD88-mediated signaling pathways and further activated TFs such as IRF7 to induce a type I IFN response. At the later time point, genes of the MyD88-mediated signaling pathway were more considerably affected by bacterial signals in comparison to the earlier time point. They now also showed increase in AP-1 expression resulting in a further increase of pro-inflammatory gene expression. As the down-stream cascade of C-type lectin receptors, the Syk-CARD9 signaling molecules were slightly regulated on transcriptional level, which probably leads to increased expression of NFκB (Marakalala et al., 2010). Cal might also enter the cytoplasm, which might trigger the MDA5-MAVS intracellular receptor and the signal transduction gene TBK-1 in macrophages to initiate a type I IFN response (Jaeger et al., 2015). Collectively, the protective effects of human PBMC against bacteria and fungi undergo a very dynamic complex regulatory mechanism involving PRRs, signaling transduction genes, TFs and other co-regulators.

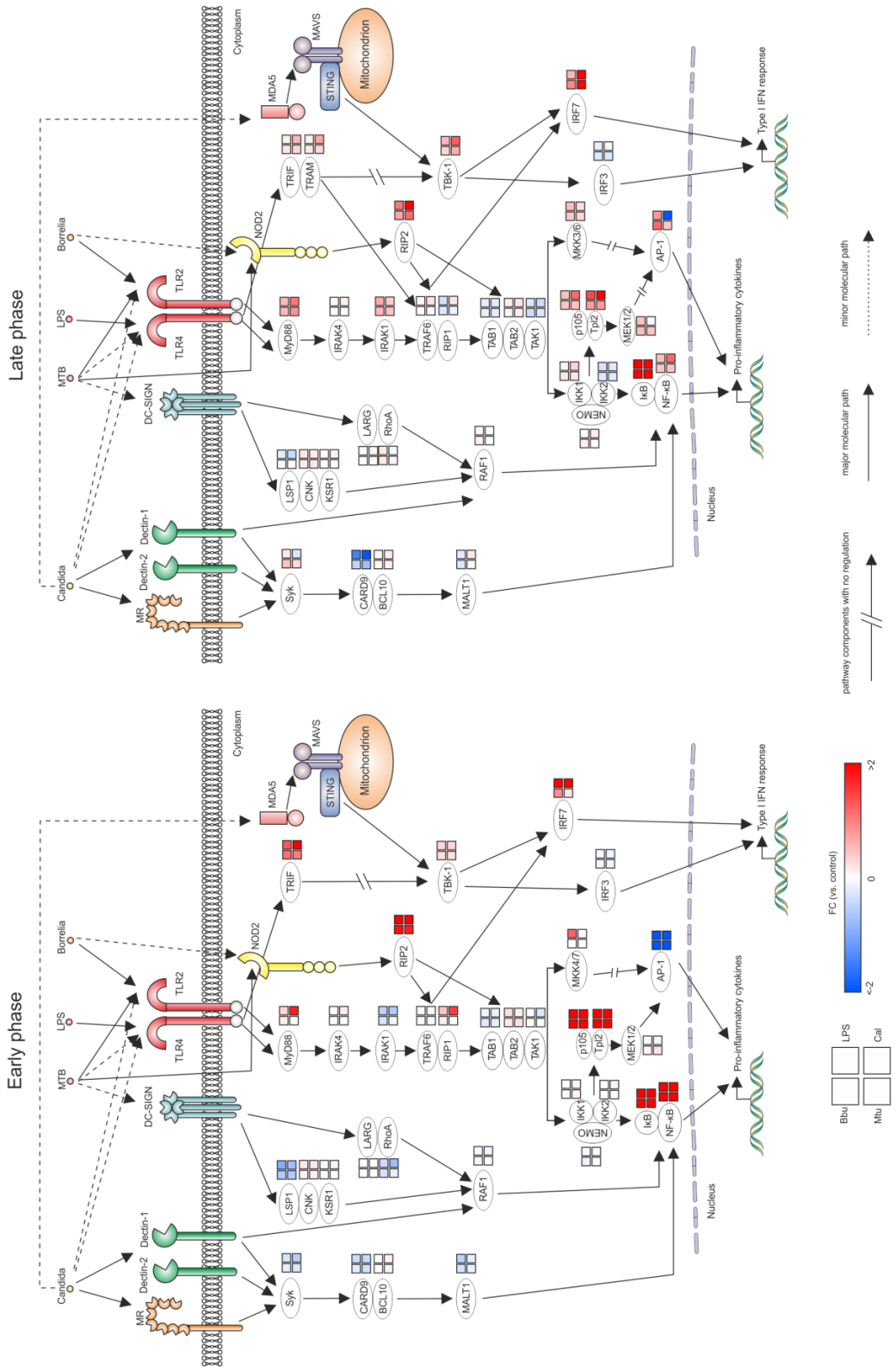

Figure 6.6 Model of regulatory pathways of bacterial and fungal stimulation. 4 and 24 hour time points are separately shown as "Early phase" and "Late phase", respectively. Each pathway component is coupled with the expression changes in 4 conditions represented by heatmap from red (FC > 2) via white (no change) to blue (FC < -2). The pathway genes with no regulation at all were removed. The solid lines stand for major molecular paths while the dashed lines are minor molecular paths.

In summary, transcriptional network analysis revealed common and stimulus-specific activation programs during the host defense against bacterial and fungal pathogens by PBMC.

7 Origin of murine tissue macrophages and the role of Irf8 in tissue macrophage subsets

Three major subjects in macrophage biology research are currently the ontogeny of macrophages during embryogenesis, the transcriptional regulation of tissue macrophage programming and the description of transcription factors involved in these processes. Since these aspects cannot be studied in humans, respective animal models have to be utilized. I had the great opportunity to address these important issues of macrophage biology in collaboration between our department and the group of Prof. Marco Prinz at the University of Freiburg. In **Chapter 7**, I describe my assessment of transcriptional changes in embryonic and adult tissue macrophages isolated from different tissues in either wild type ($IRF8^{+/+}$) or IRF8 knockout mice ($IRF8^{-/-}$).

7.1 Embryonic tissue macrophage development from yolk sac

Diverse cell populations were isolated from embryonic and adult mice by FACS sorting. To understand the influence of Irf8 on macrophage development and function, cells from Irf8 deficient mice were also derived. A total number of 27 cell populations were used for transcriptional expression profiling as it is described in **Table 7.1**.

To assess the genome-wide sample-to-sample relationships and verify the quality of the transcriptomic data, 17,888 present genes (at least one cell population is over the background) were extracted from the normalized and batch corrected data. Based on the filtered informative genes, several bioinformatics approaches were performed on yolk sac macrophages (EMP, A1 and A2) and embryonic $F4/80^{high}$ tissue cells (Emb_F4/80_kidney, Emb_KC and Emb_MG) to demonstrate if $F4/80^{high}$ cells are derived from yolk sac as displayed in the diagram **Figure 7.1.1**. PCA, co-regulation analysis and hierarchical clustering of these transcriptomes suggested a model of development of macrophages starting from EMP, via the A1 and A2 stages toward the end-differentiated cell types identified within the three tissues kidney, liver and central nervous system (CNS). Overall, transcriptional profiling clearly supported the hypothesis that yolk sac macrophages are the origin of $F4/80^{high}$ cells (**Figure 7.1.2-7.1.4**).

Table 7.1 Overview of gene expression profiles from 27 cell populations

Cell population	Organ	Time point	Primitive/Definitive Hematopoiesis	Number of arrays
EMP	Yolk sac	dpc 8.0	Primitive	3
A1	Yolk sac	dpc 9.0	Primitive	3
A1_Irf8$^{-/-}$	Yolk sac	dpc 9.0	Primitive	3
A2	Yolk sac	dpc 9.0	Primitive	3
A2_Irf8$^{-/-}$	Yolk sac	dpc 9.0	Primitive	4
Emb_MG	CNS	dpc 14.0	Primitive	3
Emb_MG_Irf8$^{-/-}$	CNS	dpc 14.0	Primitive	3
Emb_KC	Liver	dpc 14.0	Primitive	3
Emb_KC_Irf8$^{-/-}$	Liver	dpc 14.0	Primitive	3
Emb_CD11b_liver	Liver	dpc 14.0	Definitive	3
Emb_CD11b_liver_Irf8$^{-/-}$	Liver	dpc 14.0	Definitive	3
Emb_F4/80_kidney	Kidney	dpc 14.0	mixed	2
Emb_F4/80_kidney_Irf8$^{-/-}$	Kidney	dpc 14.0	mixed	3
Emb_CD11b_kidney	Kidney	dpc 14.0	Definitive	3
Emb_CD11b_kidney_Irf8$^{-/-}$	Kidney	dpc 14.0	Definitive	3
A_MG	CNS	P60	Primitive	3
A_MG_Irf8$^{-/-}$	CNS	P60	Primitive	3
A_KC	Liver	P60	Primitive	3
A_KC_Irf8$^{-/-}$	Liver	P60	Primitive	3
A_CD11b_liver	Liver	P60	Definitive	3
A_CD11b_liver_Irf8$^{-/-}$	Liver	P60	Definitive	3
A_F4/80_kidney	Kidney	P60	mixed	3
A_F4/80_kidney_Irf8$^{-/-}$	Kidney	P60	mixed	3
A_CD11b_kidney	Kidney	P60	Definitive	3
A_CD11b_kidney_Irf8$^{-/-}$	Kidney	P60	Definitive	3
A_LC	Skin	P60	Definitive	3
A_LC_Irf8$^{-/-}$	Skin	P60	Definitive	3

Abbreviations: EMP: erythromyeloid progenitor; A1: early myeloid progenitor; A2: immature yolk sac macrophage; Emb_MG: embryonic microglia; Emb_KC: embryonic Kupffer cells; Emb_CD11b_liver: embryonic CD11b^{high} macrophages from liver; Emb_F4/80_kidney: embryonic F4/80^{high} macrophages from kidney; Emb_CD11b_kidney: embryonic CD11b^{high} macrophages from kidney; A_MG: adult microglia; A_KC: adult Kupffer cells; A_CD11b_liver: adult CD11b^{high} macrophages from liver; A_F4/80_kidney: adult F4/80^{high} macrophages from kidney; A_CD11b_kidney: adult CD11b^{high} macrophages from kidney; CNS: central nervous system; dpc: days post coitum.

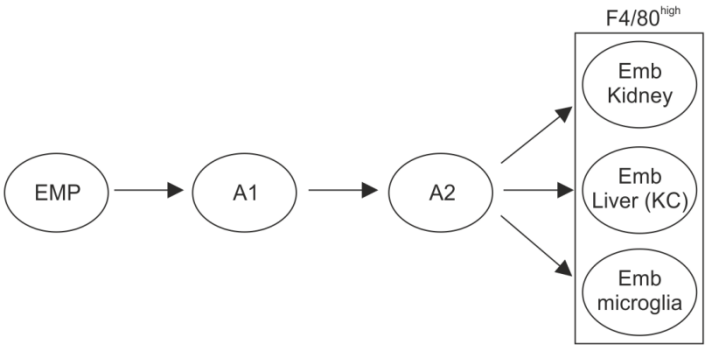

Figure 7.1.1 Schema of yolk sac macrophage development to F4/80^{high} tissue cells. Abbreviations see **Table 7.1**.

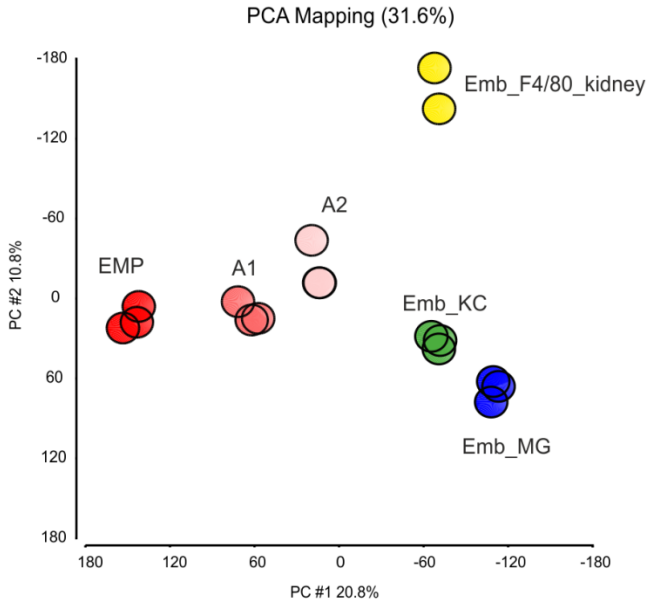

Figure 7.1.2 Principal component analysis (PCA) of embryonic macrophage development. Only first two principal components of the samples are shown. Abbreviations see **Table 7.1**.

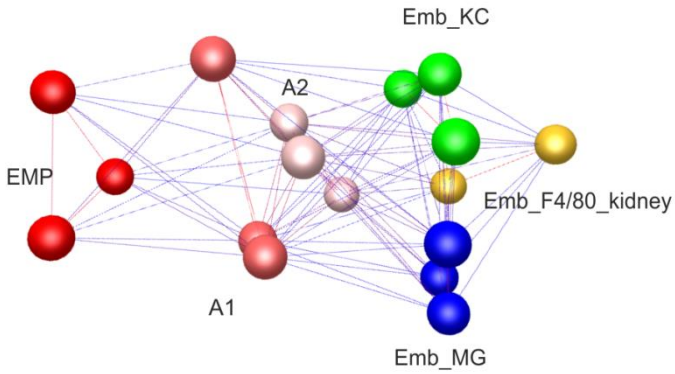

Figure 7.1.3 Sample-sample co-regulation network of embryonic macrophage development. Pearson correlation threshold of 0.92 was used. Abbreviations see **Table 7.1**.

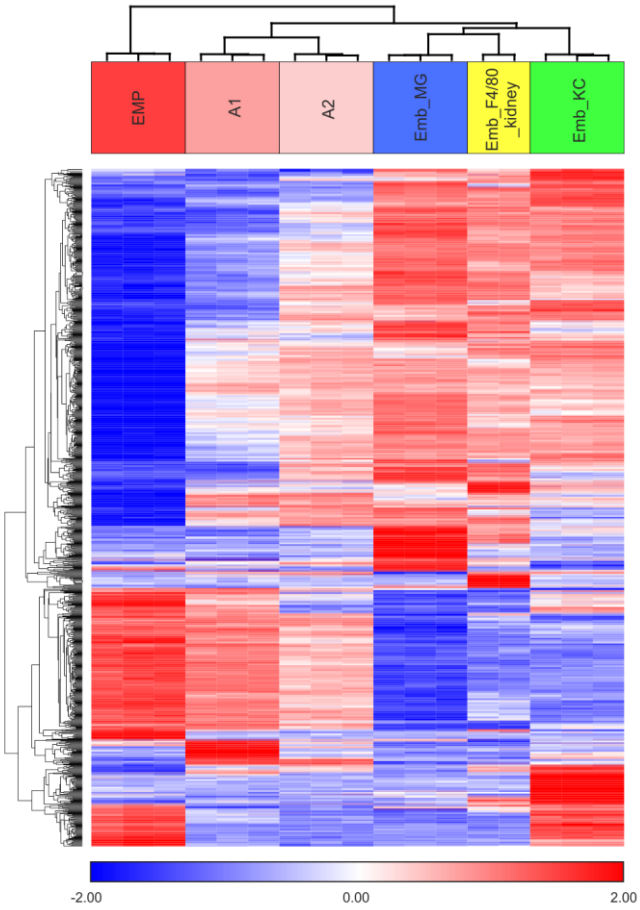

Figure 7.1.4 Hierarchical clustering of embryonic macrophage development based on 1,000 most variable probes. Heatmap displays z-transformed expression values (z-scores) from blue to red via white. Abbreviations see **Table 7.1**.

By K-means clustering on genes expressed in at least one of these cell populations, gene clusters with similar expression patterns throughout developmental stages were identified. According to Davies-Bouldin score testing (Davies and Bouldin, 1979), the complete transcriptome was partitioned into 36 clusters. Gene clusters characterized by increased expression (Cluster 3) and decreased expression (Cluster14) with later time points in development were determined. In order to address the biological processes they are involved, GO enrichment analysis was performed on clusters 3 and 14 separately. Genes related to immune system process, antigen processing and presentation, myeloid cell differentiation were up-regulated over developmental stages while genes associated with cell cycle, DNA repair and damage were found to be repressed, further supporting that tissue macrophages gain their specific functions, while losing proliferative capacities typically found in stem cells and progenitor cells (**Figure 7.1.5-7.1.6**).

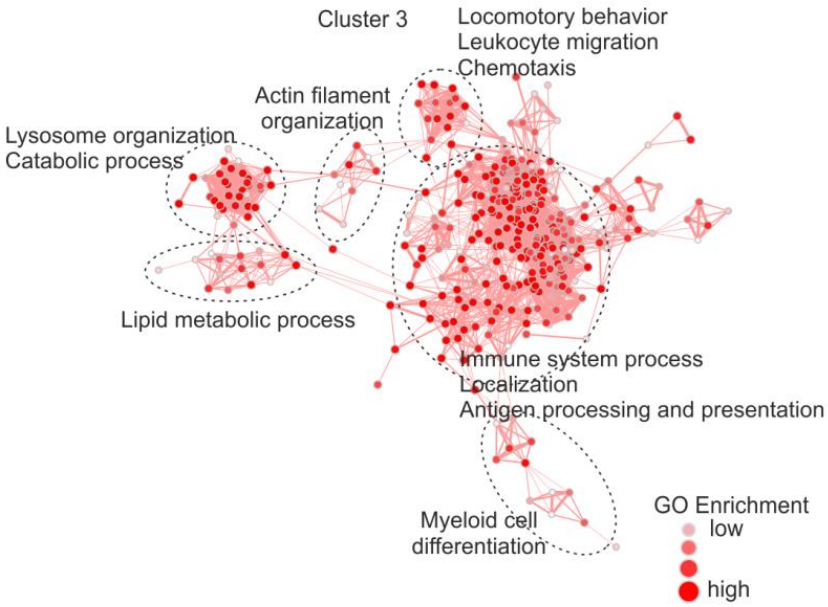

Figure 7.1.5 GO network on Cluster 3 genes identified by K-means clustering. Node size and node color darkness correspond with the enrichment FDR adjusted p-value (q-value) of the GO term, i.e. the biggest and darkest nodes are the GO terms most significantly enriched with lowest FDR q-value. Edge thickness shows overlap of genes between neighbor nodes.

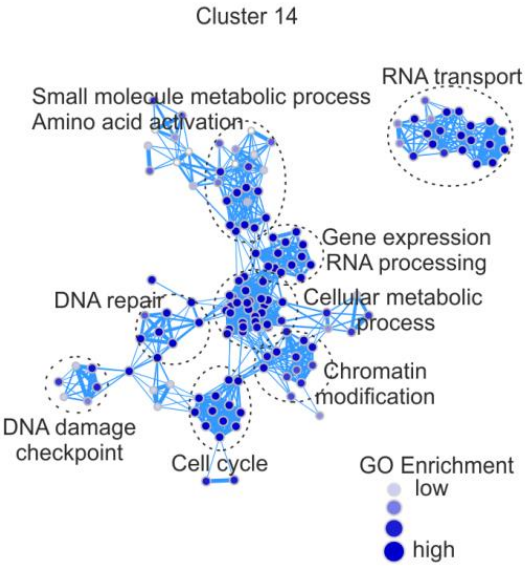

Figure 7.1.6 GO network on Cluster 14 genes identified by K-means clustering. Node size and node color darkness correspond with the enrichment FDR adjusted p-value (q-value) of the GO term, i.e. the biggest and darkest nodes are the GO terms most significantly enriched with lowest FDR q-value. Edge thickness shows overlap of genes between neighbor nodes.

7.2 Distinct transcriptional profiles of F4/80^{high} and CD11b^{high} tissue macrophages

There have been controversies about the use of different markers to isolate cells for the assessment of certain macrophage populations (Hume and Freeman, 2014; Hume et al., 2013). This issue was addressed here by analyzing both F4/80^{high} as well as CD11b^{high} cells individually. To compare the F4/80^{high} and CD11b^{high} populations systematically, two approaches were combined: one-way ANOVA and MCL clustering on a co-regulation network to identify differentially expressed (DE) genes between F4/80^{high} (MG, KC and F4/80_kidney) and CD11b^{high} (CD11b_liver and CD11b_kidney) groups. **Figure 7.2.1** and **Figure 7.2.2** illustrate the hierarchical clustering of the combined DE genes between the two groups from embryonic and adult cells, respectively. Irrespective of variation across organs, F4/80^{high} and CD11b^{high} macrophages have distinct transcriptional profiles.

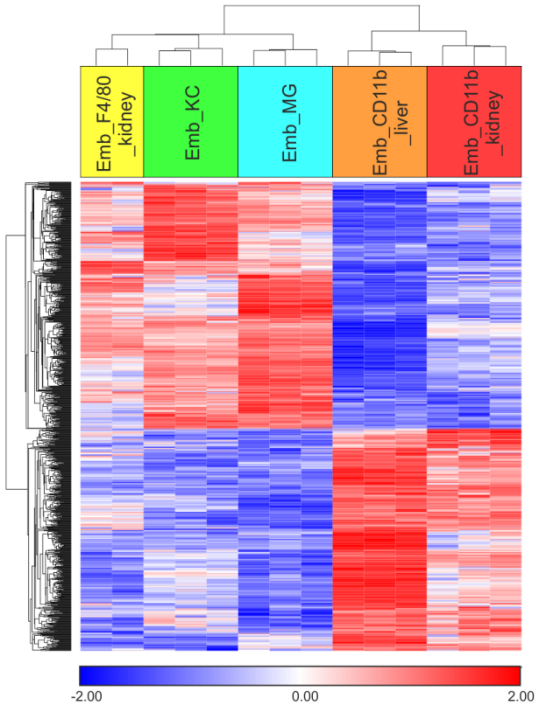

Figure 7.2.1 Hierarchical clustering of embryonic F4/80^{high} and CD11b^{high} cells based on differentially expressed genes (n = 1010). Heatmap displays z-transformed expression values (z-scores) from blue to red via white. Abbreviations see **Table 7.1**.

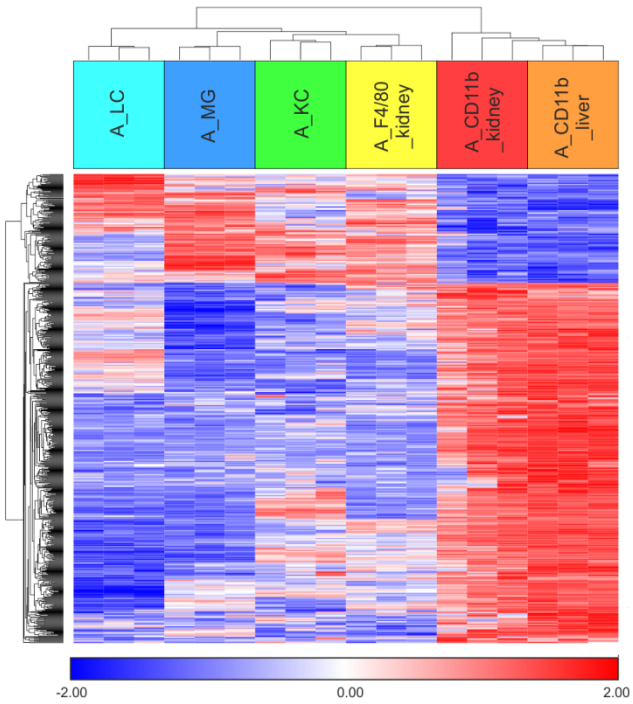

Figure 7.2.2 Hierarchical clustering of adult F4/80^{high} and CD11b^{high} cells based on differentially expressed genes (n = 755). Heatmap displays z-transformed expression values (z-scores) from blue to red via white. Abbreviations see **Table 7.1**.

In order to dismiss any impact of tissue origin, a more stringent filtering procedure was used (**Figure 7.2.3**) to define commonly expressed genes in either F4/80^{high} or CD11b^{high} cells. These non-differentially expressed (non-DE) genes throughout organs were determined by t-test over the samples within one group (FDR adjusted p-value > 0.95). By grouping F4/80^{high} (MG, liver and kidney) and CD11b^{high} (liver and kidney) populations (**Figure 7.2.3A**), 336 and 631 non-DE genes in F4/80^{high} and CD11b^{high} cells, respectively, were identified for embryonic macrophages (**Figure 7.2.3B**, left panel), and subsequently 43 DE (26 up-regulated and 17 down-regulated) genes between the two groups (irrespective of tissue of origin) were determined (**Figure 7.2.3C**, left panel). Similarly, 117 up-regulated genes and 92 down-regulated genes were identified by comparing F4/80^{high} and CD11b^{high} populations for adult cells (**Figure 7.2.3B-C**, right panel), indicating that the difference between the two enlarged during adulthood.

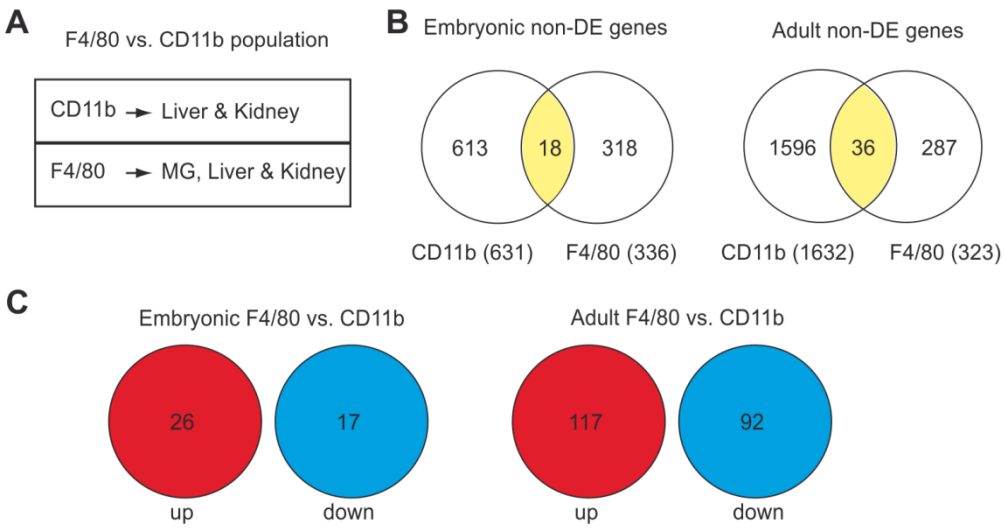

Figure 7.2.3 Differentially expressed (DE) genes between F4/80^{high} and CD11b^{high} cells irrespective of organs. A Diagram illustrates the grouping of organ-specific F4/80^{high} (MG, liver and kidney) and CD11b^{high} (liver and kidney) populations for ANOVA calculation. **B** Venn diagrams overlaying the identified non-DE genes within each group (FDR adjusted p-value > 0.95) for both embryo and adult cells in ANOVA model. **C** The number of up- and down-regulated genes in F4/80^{high} versus CD11b^{high} populations (|FC| > 1.5, FDR adjusted p-value < 0.05) from genes fulfilling the criteria in **B**.

As it is still elusive how adult Langerhans cells (LC) are developing and which myeloid compartment they belong to, their similarities and dissimilarities to both F4/80^{high} and CD11b^{high} populations were studied. Using a one-way ANOVA model, 642 DE genes between LC and F4/80^{high} and 526 DE genes between LC and CD11b^{high} cells were identified (**Figure 7.2.4**). Among them, 415 were common DE genes to both groups, which indicated that LC are very distinct from other tissue macrophages.

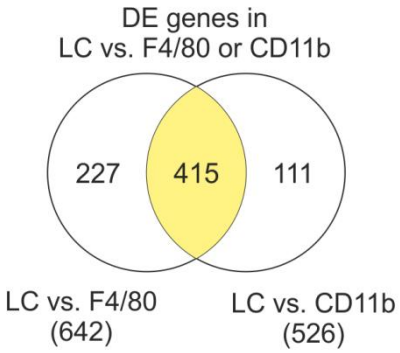

Figure 7.2.4 Comparison of adult Langerhans cells (LC) with F4/80^{high} and CD11b^{high} cells. Venn diagram shows the differentially expressed genes between adult LC and F4/80^{high} or CD11b^{high} cells (|FC| > 1.5, FDR adjusted p-value < 0.05).

7.3 Irf8 deficiency has an impact on all tissue macrophages

Next, the question whether Irf8 deficiency changes the gene expression profiles for tissue macrophages was addressed. By comparing the cells from Irf8 deficient ($Irf8^{-/-}$) mice with their corresponding wild type ($Irf8^{+/+}$) mice, up- and down-regulated genes were observed by a certain extent in 11 cell populations. As it is shown in **Figure 7.3.1**, Irf8 deficiency had mostly changed the embryonic Kupffer cells followed by adult Langerhans cells on transcriptional level. The smallest changes were observed in both embryonic and adult $F4/80^{high}$ kidney macrophages. Collectively, this analysis indicated that Irf8 plays a crucial role in all tissue macrophages.

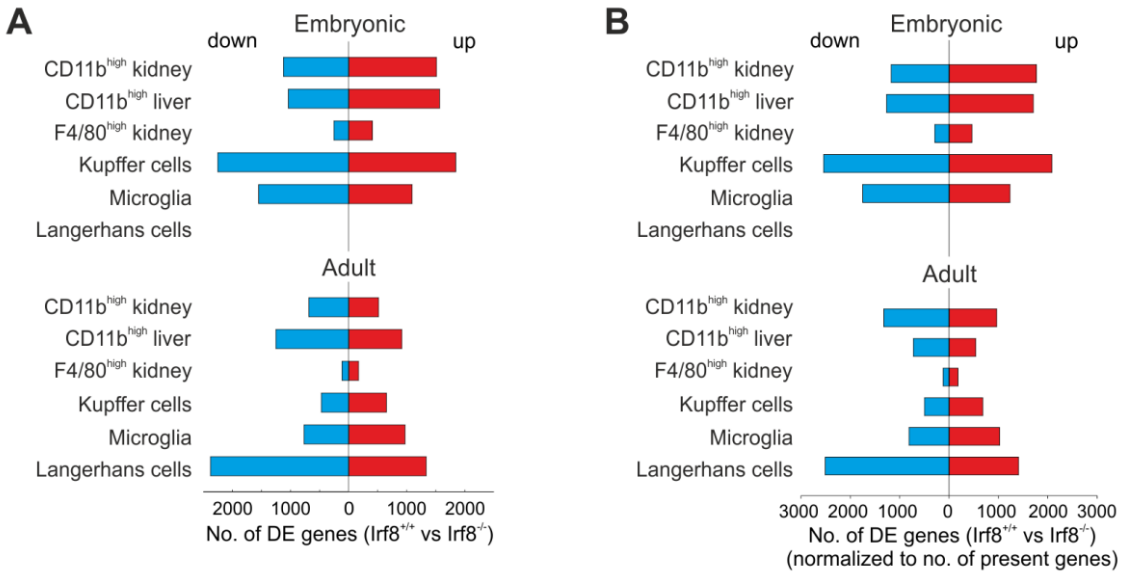

Figure 7.3.1 Number of differentially expressed (DE) genes between $Irf8^{+/+}$ and $Irf8^{-/-}$ mice in 11 groups of cell populations. DE genes were identified by one-way ANOVA model (|FC| > 1.5, FDR adjusted p-value < 0.05). **A** absolute gene numbers; **B** gene numbers normalized to the number of present genes in embryonic or adult cells.

To determine how Irf8 influences gene regulation in both $F4/80^{high}$ and $CD11b^{high}$ populations irrespective of organ difference, the differentially expressed genes ($Irf8^{+/+}$ vs. $Irf8^{-/-}$) identified in each individual cell population from different organs (liver and kidney cells for $CD11b^{high}$ group; liver, kidney and microglia for $F4/80^{high}$ group) were overlaid in Venn diagrams as it is shown in **Figure 7.3.2**. Again, in general, more transcriptional changes and also common alterations among organs were observed in embryonic cells in comparison to adult cells, indicating that Irf8 influences embryonic tissue macrophages more significantly than adult tissue macrophages.

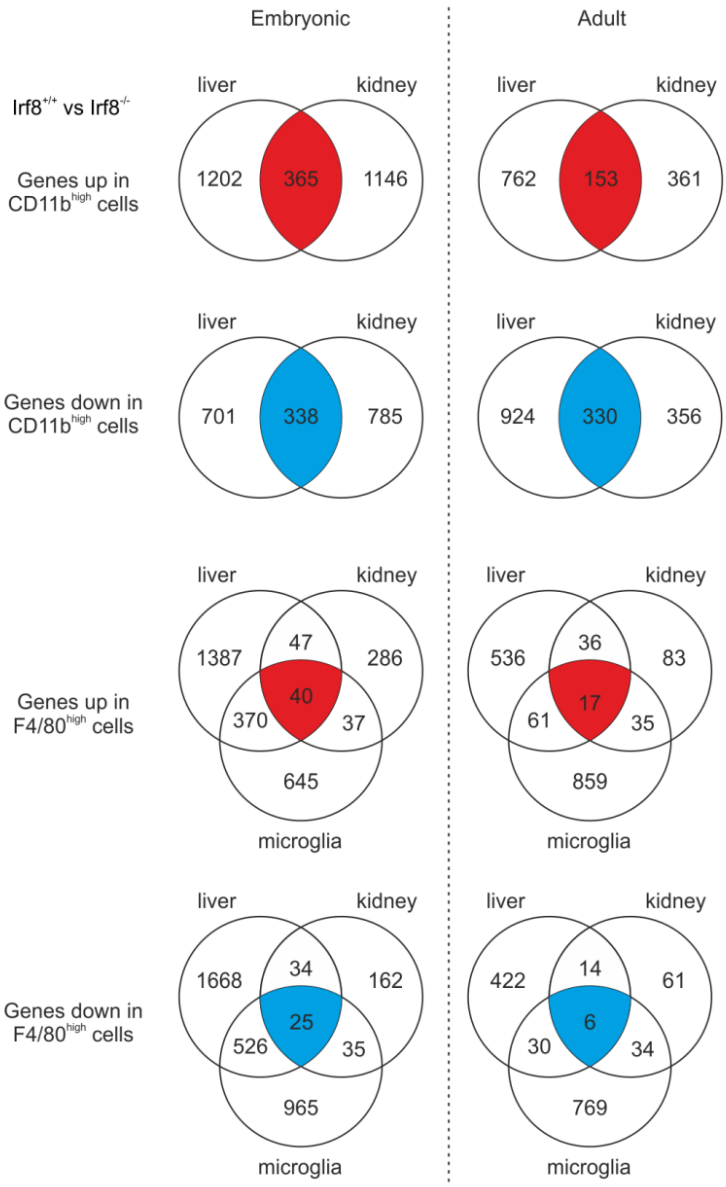

Figure 7.3.2 Venn diagrams of differentially expressed genes in F4/80^{high} and CD11b^{high} populations throughout organs. Red and blue highlighted areas are commonly up- and down-regulated, respectively in Irf8$^{+/+}$ compared to Irf8$^{-/-}$ mice.

Next, the common DE genes for F4/80^{high} and CD11b^{high} cell populations identified in at least two organs were extracted , and GO enrichment analysis was then performed to further study their known cellular functions (**Figure 7.3.3-7.3.6**). The significantly enriched GO-terms (FDR adjusted p-value < 0.05) were visualized as a network, where the GO-terms derived from up- and down-regulation were distinguished by red and blue connections (edges), respectively. The genes altered in embryonic CD11b^{high} cells derived from Irf8 deficient mice were mainly characterized by inhibition of immune system processes, cell death, and developmental processes with very few induced biological processes (**Figure 7.3.3**). In adult

cells, however, there were more up-regulated terms such as microtubule-based process, cell cycle and DNA conformation change in knock-out mice (**Figure 7.3.4**).

For embryonic F4/80^{high} cell populations, much more up-regulation was observed in Irf8 deficient mice, e.g. myeloid progenitor cell differentiation, localization and protein transport, regulation of cell proliferation (**Figure 7.3.5**). With adult cells, the functional changes caused by Irf8 deficiency became less according to very limited number of significant enriched GO terms including leukocyte differentiation and cell adhesion (**Figure 7.3.6**).

Taken together, combination of transcriptional analysis and functional genomics reflected that Irf8 plays an important role in all types of tissue macrophages, but especially triggers developmental processes in the embryo.

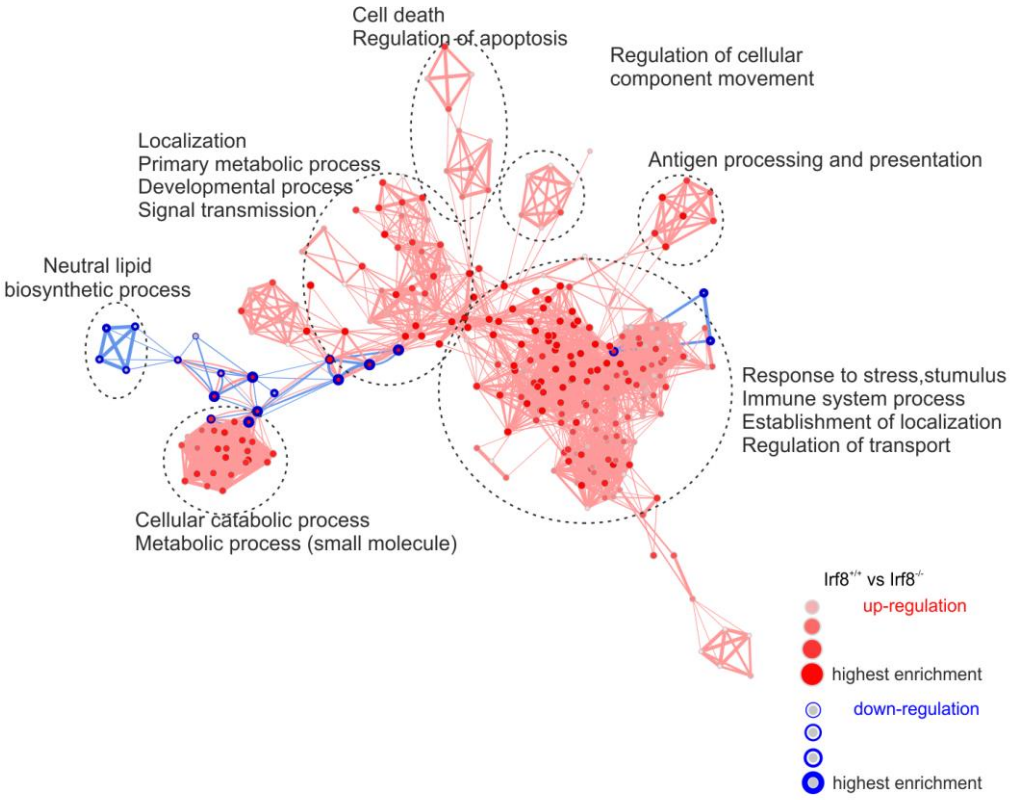

Figure 7.3.3 GO network based on common differentially expressed genes in embryonic CD11b^{high} cells derived from liver and kidney in either Irf8$^{+/+}$ or Irf8$^{-/-}$. Red edges represent GO terms associated with up-regulated genes in Irf8$^{+/+}$ compared to Irf8$^{-/-}$ mice while blue edges stand for down-regulation. Edge thickness shows overlap of genes between neighbor nodes. Node size and node color darkness (up-regulation) or node border thickness (down-regulation) correspond with the enrichment FDR adjusted p-value (q-value) of the GO term.

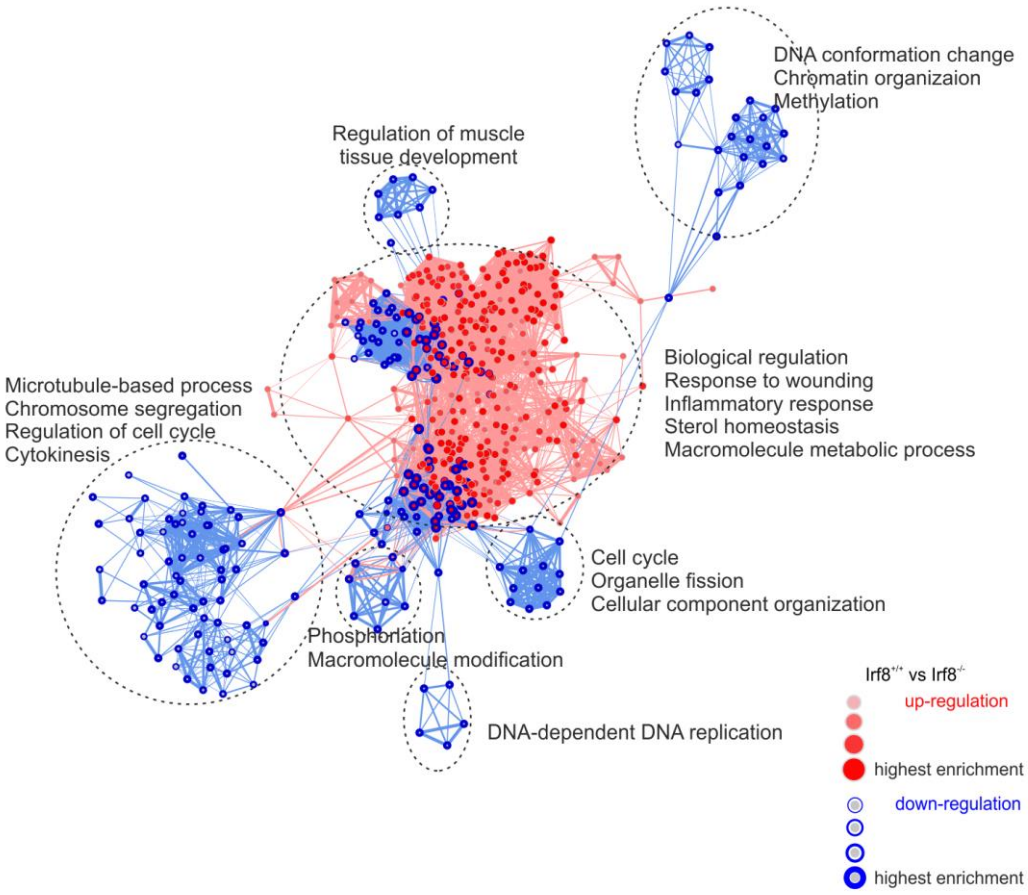

Figure 7.3.4 GO network based on common differentially expressed genes in adult CD11b^{high} cells derived from liver and kidney in either Irf8$^{+/+}$ or Irf8$^{-/-}$. Red edges represent GO terms associated with up-regulated genes in Irf8$^{+/+}$ compared to Irf8$^{-/-}$ mice while blue edges stand for down-regulation. Edge thickness shows overlap of genes between neighbor nodes. Node size and node color darkness (up-regulation) or node border thickness (down-regulation) correspond with the enrichment FDR adjusted p-value (q-value) of the GO term.

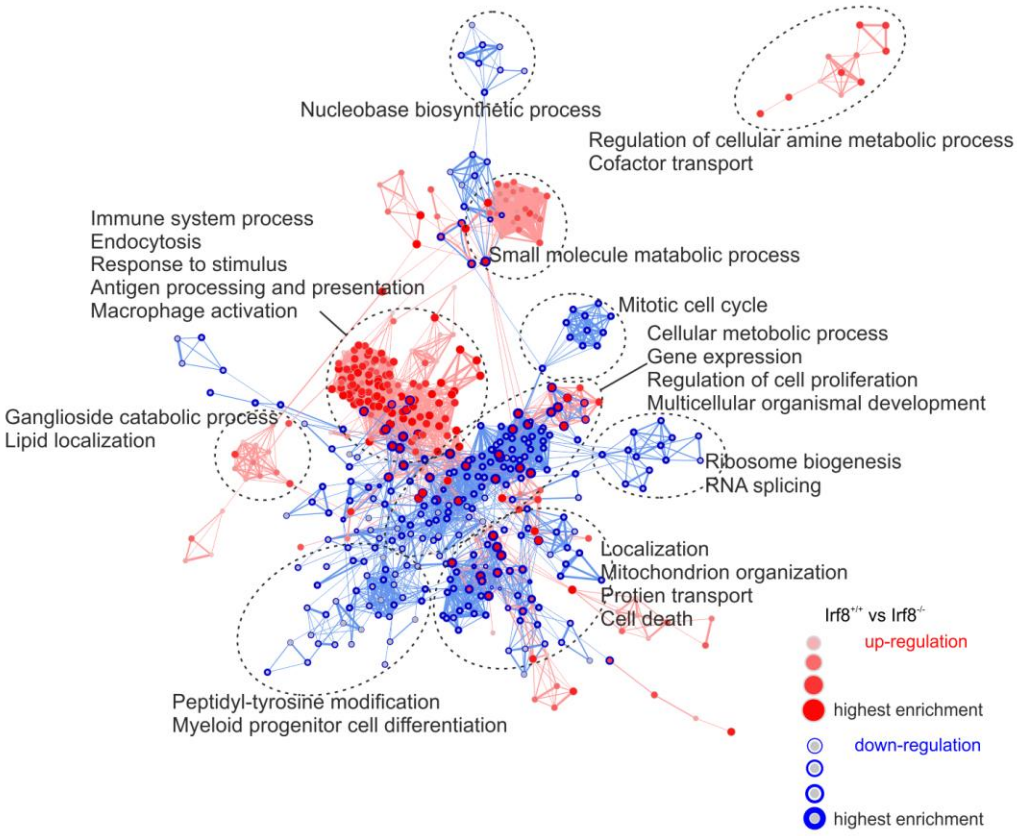

Figure 7.3.5 GO network based on common differentially expressed genes in embryonic F4/80^{high} cells derived from at least two organs in either Irf8$^{+/+}$ or Irf8$^{-/-}$. Red edges represent GO terms associated with up-regulated genes in Irf8$^{+/+}$ compared to Irf8$^{-/-}$ mice while blue edges stand for down-regulation. Edge thickness shows overlap of genes between neighbor nodes. Node size and node color darkness (up-regulation) or node border thickness (down-regulation) correspond with the enrichment FDR adjusted p-value (q-value) of the GO term.

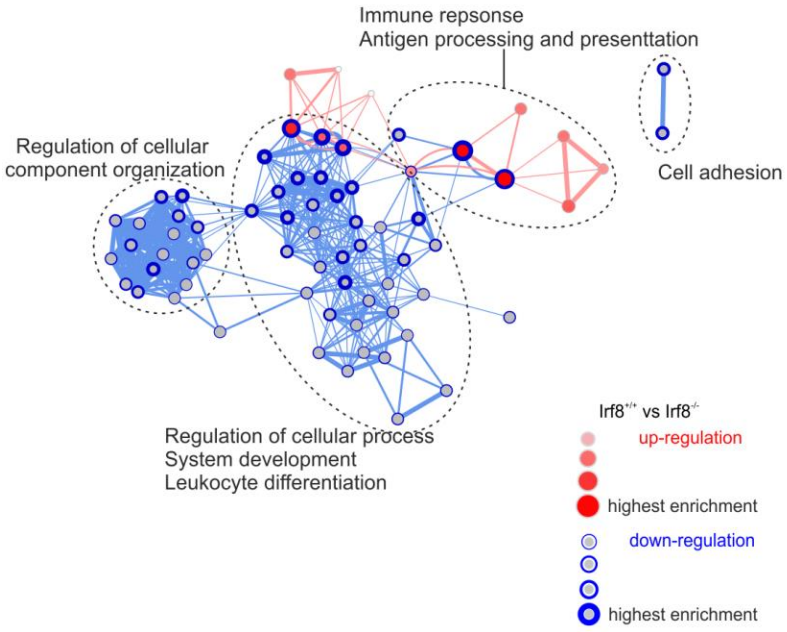

Figure 7.3.6 GO network based on common differentially expressed genes in adult F4/80^{high} cells derived from at least two organs in either Irf8$^{+/+}$ or Irf8$^{-/-}$. Red edges represent GO terms associated with up-regulated genes in Irf8$^{+/+}$ compared to Irf8$^{-/-}$ mice while blue edges stand for down-regulation. Edge thickness shows overlap of genes between neighbor nodes. Node size and node color darkness (up-regulation) or node border thickness (down-regulation) correspond with the enrichment FDR adjusted p-value (q-value) of the GO term.

In summary, computational studies on gene expression profiling data for tissue macrophages from the embryo and adult mice demonstrated the distinct origins and transcriptional profiles of F4/80^{high} and CD11b^{high} tissue macrophage populations. Global comparison of macrophages from wild type and Irf8-deficient mice indicated that Irf8 plays a crucial role in tissue macrophage development.

8 Discussion and future perspectives

In four individual studies, I have applied computational methods to better understand transcriptional regulation as an important biological mechanism in macrophage biology. In the first part, I systematically analyzed a large gene expression dataset generated from 29 human macrophage *in vitro* conditions to assess transcriptional regulation during human macrophage activation by comparing a diverse set of stimuli on a single microarray platform under highly standardized conditions. Network modeling of this dataset revealed an extension of the current M1 versus M2 polarization model into a "multi-dimensional model" with at least 9 distinct macrophage activation programs. Applying these transcriptional programs to human *in vivo* alveolar macrophages from smokers and patients with chronic obstructive pulmonary disease (COPD) revealed an unexpected biology. Reverse network engineering of the complete dataset using single inference method results in common regulatory denominators such as NFKB1.

Next, I refined the common gene regulatory network of human macrophage activation by integrating multiple network inference approaches as a community-based method. The community-based method sharpens the resolution of the common macrophage activation regulatory networks and the refined network indicated that transcription factors are the most important components in regulatory circuits involved in macrophage activation.

To assess the differentiation of human monocytes to macrophages in a complex cellular system, I analyzed a large gene expression dataset generated from infected human PBMC in a collaboration with the group of Prof. Mihai G. Netea in Nijmegen. By applying the same computational methodologies to the PBMC, I identified common and stimulus-specific transcriptional programs in host defense against bacteria and fungi. Combination of knowledge-based and data-driven analysis revealed refined pathway models for these microbial infections on transcriptional level.

Finally, in a collaboration with the group of Prof. Macro Prinz in Freiburg, I performed computational analysis on gene expression profiles for embryonic and adult tissue macrophages derived from wild type and Irf8-deficient mice. This study revealed the distinct origins and transcriptional profiles of different tissue macrophage subpopulations and a crucial role of Irf8 in development.

In the following paragraphs, I will discuss important points that resulted from my studies here in more detail.

8.1 Multi-dimensional model of macrophage activation expands the M1 versus M2 dichotomous system

Cells from the immune system show high diversity and plasticity in respect to their maturations status and local environment. Data from **Chapter 4** demonstrate that under inflammatory conditions macrophages do not always polarize into two extreme states (M1 versus M2) but can drive towards diverse directions on both transcriptional level and functional level in response to different environmental signals. For instance, a triple stimulus TPP induces a STAT4-associated transcriptional program normally observed in granulomatous diseases. To decipher central activation and specific activation programs of human macrophages, I applied different but also complementary approaches of network analysis ranging from co-regulation analysis by BioLayout, weighted gene co-expression network analysis (WGCNA) defining specific differentiation programs, to reverse network engineering using ARACNe or TINGe for decoding the major activation program. Importantly, I was able to validate findings from network analysis with additional approaches including SOM clustering, correlation coefficient matrices, hierarchical clustering or PCA as shown for the delineation of macrophages from other control cells or the development of the multiple activation states from baseline macrophages (**Chapter 4.1** and data not shown). Furthermore, it also significantly extended our understanding of these processes by developing macrophage polarization into a multi-dimensional model, by defining novel gene regulatory networks (**Chapter 4.6** and **Chapter 5**) or by prioritizing cell surface markers distinguishing macrophages from other myeloid cells both in humans and mice (Xue et al., 2014).

Linking transcriptomes derived from *in vitro* macrophage models of activation to transcriptional reprogramming in human tissue macrophages is another desired direction of this research (**Chapter 4.5**). This field is currently hindered by technical limitations, such as the difficulty in gaining sufficient numbers of highly purified macrophages from inflamed or otherwise disturbed human tissues. In fact, some of the few studies so far have focused on alveolar macrophages obtained by bronchoalveolar lavage (Shaykhiev et al., 2009; Woodruff et al., 2005). Combining these datasets and linking the *in vitro* differentiation programs via GSEA to the *in vivo* signatures revealed novel and unexpected results for alveolar

macrophages from smokers demonstrating a rather classical M2-like signature combined with loss of signs of inflammation induced by TLR ligands, TNF, IFNγ, and prostaglandins. COPD, one of the potential pathophysiological outcomes of chronic smoking exposure was still characterized by lack of an acute inflammatory response to TLR ligands, TNF and prostaglandins in alveolar macrophages, yet the IL4/IL13-driven response was no longer present. However, in contrast to previous interpretations of alveolar macrophage activation in COPD which suggest a mixed M1- and M2-phenotpye (Houghton, 2013), I unequivocally demonstrate by the data-driven analysis that alveolar macrophages in COPD are neither M1- nor M2-activated. This can be easily explained by the fact that many of the marker genes used in prior studies are not specific to IFNγ- or IL4-mediated reprogramming, but are rather induced by a larger set of different stimuli. The lack of any enrichment of the 49 gene sets derived from the *in vitro* macrophage activation model in alveolar macrophages from COPD patients might be explained by a reduced capability of alveolar macrophages to gain any immune functions during fibrotic remodeling of lung tissue in COPD. An alternative explanation might be that factors not assessed in our model but linked to COPD pathophysiology including TGFβ, VEGF, angiostatic peptides (angiostatin, endostatin) or extracellular matrix substrates might be major factors reprogramming alveolar macrophages *in vivo* (Houghton, 2013). A dominant role for such 'non-classical' macrophage activation signals might be further supported by clinical findings that classical anti-inflammatory therapy provides little or no benefit to COPD patients (Cazzola et al., 2012). Overall, this data-driven approach can contribute to the understanding of biological processes operative in alveolar macrophages *in vivo* and these findings further support the proposed multi-dimensional model of macrophage activation.

While *in vivo* validation studies using genetic mouse models have clearly indicated that TFs such as STAT1 and STAT6 are linked to IFNγ- and IL4-induced macrophage differentiation programs, respectively (Lawrence and Natoli, 2011; Spence et al., 2013), several of the TFs identified here are novel and appear be part of larger TF networks. They have not been studied in the context of either activation or differentiation of macrophages. In fact, for many of these genes the respective tools ranging from specific antibodies for biochemical and epigenomic analysis as well as suitable genetic models are currently unavailable. Technological advances such as the generation of tagged proteins combined with high-throughput ChIP-Seq (Blecher-Gonen et al., 2013), and the emergence of gene editing

approaches including TALENs (Bedell et al., 2012), and CRISPR/Cas systems (Cong et al., 2013) will most likely make it possible to elucidate the role of individual TFs during complex processes such as macrophage activation or differentiation.

Macrophages are very important players in maintaining many aspects of tissue homeostasis and mediating various acute inflammatory reactions as a host response to endogenous or exogenous stress signals. Also in chronic inflammatory conditions, macrophages are probably the most prominent cells of the myeloid compartment. However, their impact on many of the common diseases such as cancer, obesity, atherosclerosis, autoimmune disorders is still underappreciated. Thus, it will be tremendously important to understand how macrophages gain their functional repertoire, how they can change their function and their localization for better diagnosis and treatments of these diseases. Since the activation and function of macrophages are tightly dependent on transcriptional regulation, genome-wide transcriptional studies at both cell population and single cell level will be still one of the major directions to address current and future questions of human macrophage biology in homeostasis and disease. The conservation of expression between human and animal models is still obscure, and therefore it might be wiser to study macrophage biology directly in humans, rather than focusing on animal models (Davis, 2008). A prerequisite for such an effort will be to access to not only *in vitro* generated macrophages but also the assessment of human tissue macrophages. To achieve this, research networks and consortia will have to be formed because single laboratories would be insufficient (Schultze et al., 2015). Higher resolution data from genomic technologies such as next-generation sequencing will become more and more prominent to study human macrophage biology in homeostasis and diseases. Although current data highlight information on the level of tissues or defined cell types, spatio-temporal information about signal integration and transcriptional regulation of human macrophages are still missing. Hence, systems immunology will be a key aspect in the future, and more robust tools and technologies need to be developed to study such questions with minimum effort and very little liability on our patients.

8.2 Community-based methods outperform individual methods

To improve our mathematical modelling of macrophage activation, I determined the impact of a DREAM-based model on the description of macrophage activation by utilizing our previously reported dataset of macrophage transcriptomes from diverse signal inputs,

applying five network inference methods and integrating the individual predictions to form a consensus network (**Chapter 5**).

Integrating predictions from different, especially complementary methods performed better than using individual methods. For instance, according to the assessment of degree node distribution, the consensus network (**Figure 5.1.2**) has much higher R-squared correlation to the curve compared to the networks generated by single methods (**Figure 4.6.2**), suggesting that the former better fits the power law as a topologically scale-free network. Moreover, it was demonstrated by Marbach *et al.* that using complementary methods enables us to avoid method-specific biases in predicting various network motifs (Marbach et al., 2012). There is no single best method that fits all different biological, technical and experimental design conditions best: the more best-performing methods are combined, the more accurate and precise results would be obtained. Thus, it could be one of the future directions to find out better combinations of approaches for modelling of different biological systems.

The consensus network indicates that transcription factors are the most important components of macrophage activation (**Chapter 5.2**). Compared to the one-method-based network, a big increase of weightiness of transcription regulators has been observed in the community-based network. Interestingly, all the predictions from five applied methods are on the basis of all-to-all (the complete transcriptome) calculations, i.e. none of the applied methods is TF-biased. Hence, this observation should be less dependent on which individual prediction algorithms are used but mostly relies on how many robust methods are combined. Among the most prominent TFs associated with macrophage activation independent of input signal were CREB1 and BLZF1. CREB1, a member of the leucine zipper family of DNA binding proteins, is a key transcription regulator in diverse cellular responses such as cell proliferation, survival and differentiation. It is induced by a variety of growth factors and inflammatory signals and subsequently mediates the transcription of immune-associated genes containing a cAMP-responsive element, for example, IL2, IL6, IL10 and TNF. Furthermore, CREB1 induces an anti-apoptotic effect in monocytes and macrophages (Wen et al., 2010). It has been identified to interact with more than 130 proteins in previous studies (data from NCBI Gene), suggesting its importance in cellular mechanisms. BLZF1 (basic leucine zipper nuclear factor 1), however, is not studied as well as CREB1, suggesting

that transcriptome-based network analysis enables the discovery of important novel regulations during macrophage activation.

Among the most central hubs, there are genes maintaining fundamental cellular and molecular functions. e.g. INIP (INTS3 and NABP interacting protein, or C9orf80) is important for the maintenance of genome stability (Huang et al., 2009), GJC1 is involved in gap junction channel activity (Kanter et al., 1994) and the gene is silenced in cancer (Sirnes et al., 2011). Albeit not linked to immune function, the roles of these genes in common macrophage activation networks should not be disregarded.

Another important finding was the identification of distinct network clusters that reflected very specific cellular functions, which was not revealed in such clarity by individual methods (**Chapter 5.3**). This result indicates that the community-based method is more advanced in keeping more information (functional network modules) from the expression data than individual methods do. This could be explained by the limitations of individual methods to predict different network motifs. In contrast, the integration of complementary methods seems to cope much better with drawbacks of individual methods (Marbach et al., 2012).

Taken together, reverse engineering of large enough transcriptional datasets by integration of multiple inference methods sharpens the resolution of the regulatory circuits involved in human macrophage activation thereby enhancing our capacity to prioritize future validation experiments.

8.3 Refined pathway models for host defense against bacteria and fungi

PBMC are a heterogeneous population of cells including lymphocytes (T cells, B cells, and NK cells) and myeloid cells (i.e. monocytes, macrophages and DCs). In the response to infections they act in concert. A big advantage of using PBMC is that the inter-cellular interactions between these cells are also taken into account. Previous studies on bacterial infections indicated that monocytes, dendritic cells, macrophages, NK cells, NK-T cells, and polymorphonuclear cells, all contribute to generate a coordinated and robust response to Bbu infection (Moore et al., 2007; Salazar et al., 2003). Except for polymorphonuclear cells, PBMC contain all the cells involved in a coordinated response to Bbu (Cervantes et al., 2014). While single cell technologies need to be developed in the future to study the development of immune responses in complex cell compositions on a global scale, I asked whether

population based analysis such as those performed on transcriptional level in PBMC could reveal models of immune cell activation, involvement of certain cell types and transcriptional regulations. Indeed, an unexpectedly high content of information could be revealed from cell populations studies. The estimation of cell type abundance suggested that stimulation of PBMC by microbial pathogens up to 24 hours triggers mainly monocytes differentiation into macrophages and DCs, or even activated phenotypes of these myeloid cells (**Chapter 6.2**).

Acute inflammation by microorganisms relies primarily on activation of the myeloid compartment including monocytes, macrophages and DCs. By identifying common regulatory cascades, building common TF co-regulation network and mapping stimulation- and time-specific transcriptional information onto the network combined with pathway analysis (**Chapter 6.3**, **Chapter 6.4** and **Chapter 6.6**), I confirmed that NFκB and NFκB signaling genes are common central players in anti-bacterial and fungal immunity. Initiating the transcription of pro-inflammatory cytokines, NFκB signaling has been linked to microbial infections in many previous studies (Arimilli et al., 2007; Hayden and Ghosh, 2008; Ingman et al., 2014; Mohamed and McFadden, 2009; Schwartz et al., 2013; Shwetha et al., 2013; Wang et al., 2011). This result also confirms the importance of NFκB in common activation programs of human macrophages as one cellular component of PBMC (**Chapter 4.6**).

Considering the Cal-specific TF-network (**Chapter 6.4**), the enrichment of signal transducers and activators of transcription (STATs: STAT1/3/4) supported some earlier data that STAT1 and STAT3 deficient patients have chronic mucocutaneous candidiasis (Marodi et al., 2012). Moreover, the interferon regulatory factors (IRFs: IRF7/8/9/4) are important interferon pathway genes that have been described to be essential for protective immunity against Cal (Smeekens et al., 2013).

The investigation of relative fractions of 29 macrophage activation signatures from **Chapter 4** in the infected PBMC indicated that activation patterns are dependent on both stimuli and time of stimulation (**Chapter 6.5**). PBMC stimulated for 24 hours had increased phenotypes associated with chronic inflammation like M^{TPP} instead of acute inflammation such as $M^{sLPS+IFN\gamma}$. Surprisingly, the gain of upLPS and PGE_2+P3C signatures at 24 hours suggested a complex transcriptional regulation of macrophages. As expected, significant enrichment of IFNγ and IFNβ signatures in Cal-stimulated cells confirmed previous findings by functional genomics (Smeekens et al., 2013). Albeit several of these findings are in line with known

biology, this study might still contain some inaccuracies due to the time points at which the macrophage activation programs where measured (24-72 hours) while PBMC were assessed at 4 and 24 hours. Unpublished results from our macrophage activation dataset of macrophages assessed at earlier time points clearly revealed that transcriptional regulation after initiation of stimulation is time-dependent.

Moreover, different pathogens are usually recognized by different PRRs either on the membrane or in the cytoplasm of cells. Not only one particular PRR-induced pathway is signaling, but rather several PRRs are involved simultaneously. It can also not be ruled out, that different signals probably share several signaling pathways (**Chapter 6.6**). Another limitation might be the use of heat-killed bacteria. The immune system responds more robustly to viable microorganisms than it does to heat-killed microorganisms (Sander et al., 2011). Therefore, a limitation of this study might be that the heat-killed microorganisms that were used in this study might lead to a weaker response of the cells that might not be as easily detected.

Cellular responses to any external stimuli are dynamic processes, particularly on transcriptional level. As mentioned above, following initiation of stimulation, transcriptional waves are initiated. During such waves some genes can even show differential regulation with opposite directions at early and late phases, e.g. transcription regulator FOS, a member of the activator protein-1 (AP-1) that regulates cell proliferation, differentiation and transformation, were down-regulated 4 hours after initiation with bacterial stimuli but the expression was elevated after 24 hours, while the repression of FOS remained longer in Cal infection. These findings are in line with previous observations for infections by bacteria such as *Mycobacterium bovis* and Mtu (Magee et al., 2014) as well as Hepatitis C virus infections (Kasama et al., 2014) in humans and mice. However, longer stimulation may activate AP-1 member proteins to mediate cytokine production via NFκB (Yang et al., 1999). Although gene expression changes on transcriptional level might not reflect similar protein regulation, in general, fungal signals transduce quicker than bacterial stimuli.

In addition, to link these findings to microbial infectious and inflammatory diseases, besides transcriptional regulation, post-transcriptional regulation, epigenetic variations and genetic variations would be another important aspect to look into (Netea et al., 2012).

In spite of the limitations discussed above, the systematic transcriptional analysis of PBMC revealed many unexpected findings in a rather unexpected level of detail when applying state-of-the-art computational analysis.

8.4 Genomic pedigree of F4/80^{high} tissue macrophages and the impact of Irf8 on tissue macrophage homeostasis

In addition to the plasticity of monocyte-derived macrophages activated by stress-signals from their microenvironment, tissue-specific macrophages in homeostatic conditions are also heterogeneous populations in both humans and mice. To study the origins and genomic properties of tissue macrophages, cells were isolated from different stages of embryonic development as well as embryonic tissue macrophages in a murine model to perform gene expression profiling experiments. Genome-wide transcriptomic analysis on embryonic populations demonstrates that F4/80^{high} macrophages irrespective of organ localization (CNS, liver or kidney) originate from yolk sac (**Chapter 7.1**). Furthermore, transcriptomic landscape perfectly displays their relationships at different time points of embryonic developmental stages, i.e. from progenitors at EMP, A1 and A2 to the cells generated after organ formation. Therefore, F4/80 is an ideal surface marker for primitive (yolk sac-derived) macrophages.

On the other hand, F4/80^{low}CD11b^{high} populations, which are Myb-dependent during development and are believed to be derived from hematopoietic stem cells (HSC) by definitive hematopoiesis (Schulz et al., 2012), not only exist in the liver but also in the kidney. It has been shown by Movita *et al.* that CD45^{+}CD11c^{-}F4/80^{high}CD11b^{low} liver macrophages are large and contain multiple phagocytic vacuoles whereas CD45^{+}CD11c^{-}F4/80^{low}CD11b^{high} cells are smaller with fine granules in the cytoplasm and are able to resemble monocytes. Also, the abundance of the CD11b^{high} cells is strikingly growing during inflammation (Holt et al., 2008; Karlmark et al., 2009; Movita et al., 2012). In contrast to Kupffer cells, these cells have been observed to be less sensitive to gadolinium chloride- or clodronate liposome-mediated depletion and have a lower ability to produce ROS and reactive nitrogen species, suggesting that they have a low endocytic and enzymatic activity (He et al., 2009; Kono et al., 2002; Movita et al., 2012). In this study, I was able to demonstrate that despite tissue locations they have distinct transcriptional programs on global level when comparing to F4/80^{high} macrophages derived from yolk sac (**Chapter 7.2**). Their differences are found quite considerable in the embryo and the feature expands to the adulthood. Notably, however,

the dissimilarity across the organs is not so apparent for both F4/80^{high} and CD11b^{high} populations, suggesting that the genomic properties of tissue macrophages rely more on their origin than on their location. Among the DE genes between F4/80^{high} and CD11b^{high} populations, a number of surface markers such as Pmp22, Slc40a1, P2ry13, Abcc3, Scamp5 and Mertk are significantly highly expressed (FC > 2, FDR adjusted p-value < 0.05) in both embryonic and adult F4/80^{high} cells. Mertk has been reported as macrophage core signature gene distinguishing macrophages from DC although its expression level varies during human macrophage activation (Gautier et al., 2012; Xue et al., 2014). It plays a key role in enhanced phagocytosis coupled with Gas6 (Albert et al., 2015), which is also significantly induced in F4/80^{high} macrophages as shown in our study. Similarly, there are also many surface molecules that have comparably higher expression in CD11b^{high} subsets, e.g. Gpr141, Ccr2, Atp8b4, Olr1, Sell and Slco4c1. Furthermore, Cebpe is a highly expressed TR that interacts with several other macrophage lineage mediating TRs like PU.1 (Du et al., 2002) and Myb (Verbeek et al., 1999) to induce macrophage differentiation and functional maturation of committed granulocyte progenitor cells. Nonetheless, whether these two populations behave in a different way on functional level remains to be determined.

Langerhans cells (LC) reside in the epidermis and were originally considered to be the main antigen-presenting cells in the skin as the representative of the non-lymphoid tissue conventional DCs (cDC). However, unique properties have been observed for them. For instance, in the mouse, LC originate from yolk sac-derived myeloid precursors that are recruited to the epidermis around embryonic day 18 (Chorro et al., 2009) and from fetal liver-derived monocytes (Hoeffel et al., 2012), rather than from a pre-cDC precursor. Fully-differentiated LC have a low rate of in situ proliferation to maintain their numbers in adulthood independently of input from adult hematopoiesis. This contrasts with that of cDC, which requires a continuous input of bone marrow-derived, blood-borne precursors. Therefore, LC have a unique 'lifestyle' that differs from that of cDC and monocyte-derived DC in the dermis. They were characterized as a lineage that is related to prenatally established, non-lymphoid tissue macrophages (Satpathy et al., 2012), but different from tissue macrophages and have a migratory capability that is similar to that of non-lymphoid tissue cDC (Tamoutounour et al., 2013). Data from **Chapter 7.2** (**Figure 7.2.4**) suggest that LC are very distinct from other tissue both F4/80^{high} and CD11b^{high} macrophages on transcriptional level, which supports the previous findings that LC are genetically distinct from

hematopoietic stem cells (Schulz et al., 2012) and have dual origins: mainly from yolk sac and minor from fetal liver (Hoeffel et al., 2012).

By global comparison of Irf8 deficient mice with respective wild type mice on transcriptional level, I demonstrate that Irf8 plays an important role in all types of tissue macrophages, especially triggers developmental processes in the embryo (**Chapter 7.3**). Irf8 deficiency also causes large expression changes in adult LC although the cell numbers and morphology are not altered much. In contrast, the transcriptional variations in microglia of adult $Irf8^{-/-}$ mice resulted in decreased cell numbers and alterations of cell morphology such as decrease of dendritic length and dendritic terminal points (data not shown). These results indicate that changes at transcriptional level might not always reflect changes in cell morphology and macrophage recruitment in tissues. In more detail, I found that the TF Myb is commonly inhibited (embryonic kidney, adult kidney and liver cells) when Irf8 is absent, again demonstrating that production of $CD11b^{high}$ macrophages is Myb-dependent (Schulz et al., 2012). Similar results were observed for the Myb-interacting TF Cebpe, which is a key mediator for $CD11b^{high}$ subsets as shown above. Several surface molecules expressed in $CD11b^{high}$ populations also rely on the presence of Irf8, e.g. Slco4c1, Il28ra and Cd33, etc. Interestingly, among the common repressed genes in the presence of Irf8, Irf7 is a key regulator of type I IFN-dependent immune responses and plays a critical role in the innate immune response against viruses (Chariot, 2009; Ning et al., 2011). Speaking of $F4/80^{high}$ subsets, Irf8 deficiency seems to play a more important way in embryonic macrophage development rather than in the adult system. Runx2 is one of the repressed TFs shared by three organs (kidney, liver and MG) but only in embryonic cells. Additionally, functional analysis on DE genes in $Irf8^{-/-}$ mice revealed that Irf8 is essential not only in myeloid progenitor cell differentiation, system development and immune system processes, but also in antigen processing and presentation.

In conclusion, $F4/80^{high}$ and $CD11b^{high}$ cells are two distinct macrophage populations no matter which organ they are localized. The former is developed from yolk sac by primitive hematopoiesis into different tissues while the latter is derived from HSC by definitive hematopoiesis in a Myb-dependent manner. Irf8 mediates both procedures by cooperating with many regulators and signaling cascades.

Acknowledgements

I would like to thank Dr. Susanne Schmidt, Dr. Michael Mallmann, Dr. Andrea Nino-Castro, Dr. Dominic De Nardo and Dr. Larisa Labzin for providing samples used in this study as well as coordination of patient samples and clinical data; Dr. Marc Beyer and the colleagues from the Group of Prof. Dr. Joachim L. Schultze for performing Illumina microarray experiments. I thank Dr. Astrid Draffehn, Inga Quester, Lisa Schmidleithner and Heidi Theis for performing the wet lab validation experiments; Dr. Thomas Ulas for performing WGCNA analysis and Martina Emde for analyzing the miRNA-Seq data. These are also very important parts of this work. I thank Tom Wegner for computational and technical assistance; Thanks to all the Lab members and IT staff at the LIMES Institute, University of Bonn, for their ever willingness to help.

Next, I want to acknowledge Prof. Tom C. Freeman (the Roslin Institute and Royal School of Veterinary Studies, University of Edinburgh) for the guidance of using BioLayout software. I appreciate Prof. Eike Latz (Institute of Innate Immunity, University Hospitals, University of Bonn) for taking part in the macrophage multi-dimensionality study and giving good effort on the publication. And also thanks to Prof. Mihai G. Netea (Department of Medicine, Radboud University Nijmegen Medical Centre, The Netherlands) for the PBMC dataset and Prof. Marco Prinz, Dr. Nora Hagemeyer, Dr. Kathrin Frenzel (Institute of Neuropathology, University Hospitals of Freiburg) for the murine tissue macrophage datasets. I appreciate very much their efforts and suggestions on the work as well.

I give special thanks to my bioinformatics colleagues in Prof. Schultze's group: Dr. Thomas Ulas, Jil Sander and Kathrin Klee for good advices and for always being there and helping me with many aspects of my work.

At last but not least, I am extremely grateful for advices and discussions from my direct supervisor Prof. Dr. Joachim L. Schultze. I benefited from his expertise, continuing support and encouragement in all aspects of this study.

References

Abbas, A.K., and Lichtman, A.H. (2009). Basic immunology : functions and disorders of the immune system, 3rd edn (Philadelphia, PA: Saunders/Elsevier).

Abramson, S.L., and Gallin, J.I. (1990). IL-4 inhibits superoxide production by human mononuclear phagocytes. J Immunol *144*, 625-630.

Albert, R., Kristof, E., Zahuczky, G., Szatmari-Toth, M., Vereb, Z., Olah, B., Moe, M.C., Facsko, A., Fesus, L., and Petrovski, G. (2015). Triamcinolone regulated apopto-phagocytic gene expression patterns in the clearance of dying retinal pigment epithelial cells. A key role of Mertk in the enhanced phagocytosis. Biochimica et biophysica acta *1850*, 435-446.

Aluru, M., Zola, J., Nettleton, D., and Aluru, S. (2013). Reverse engineering and analysis of large genome-scale gene networks. Nucleic acids research *41*, e24.

Amer, A.O., and Swanson, M.S. (2002). A phagosome of one's own: a microbial guide to life in the macrophage. Current opinion in microbiology *5*, 56-61.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.

Arguello, R.J., Rodriguez Rodrigues, C., Gatti, E., and Pierre, P. (2015). Protein synthesis regulation, a pillar of strength for innate immunity? Current opinion in immunology *32*, 28-35.

Arimilli, S., Johnson, J.B., Alexander-Miller, M.A., and Parks, G.D. (2007). TLR-4 and -6 agonists reverse apoptosis and promote maturation of simian virus 5-infected human dendritic cells through NFkB-dependent pathways. Virology *365*, 144-156.

Auffray, C., Sieweke, M.H., and Geissmann, F. (2009). Blood monocytes: development, heterogeneity, and relationship with dendritic cells. Annu Rev Immunol *27*, 669-692.

Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., and Lewis, P.A. (1994). A study of the classification capabilities of neural networks using unsupervised learning: a comparison with k-means clustering. Psychometrika *59*, 509-525.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nature genetics *37*, 382-390.

Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug, R.G., 2nd, Tan, W., Penheiter, S.G., Ma, A.C., Leung, A.Y.*, et al.* (2012). In vivo genome editing using a high-efficiency TALEN system. Nature *491*, 114-118.

Berende, A., Oosting, M., Kullberg, B.J., Netea, M.G., and Joosten, L.A.B. (2010). Activation of innate host defense mechanisms by Borrelia. Eur Cytokine Netw *21*, 7-18.

Best, J.A., Blair, D.A., Knell, J., Yang, E., Mayya, V., Doedens, A., Dustin, M.L., Goldrath, A.W., Immunological Genome Project, C., Monach, P.*, et al.* (2013). Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. Nature immunology *14*, 404-412.

Beutler, B., and Rietschel, E.T. (2003). Innate immune sensing and its roots: the story of endotoxin. Nature reviews. Immunology *3*, 169-176.

Bezdek, J.C., and Nikhil, R.P. (1995). An index of topological preservation for feature extraction. Pattern Recognition *28*, 381-391.

Bezman, N.A., Kim, C.C., Sun, J.C., Min-Oo, G., Hendricks, D.W., Kamimura, Y., Best, J.A., Goldrath, A.W., and Lanier, L.L. (2012). Molecular definition of the identity and activation of natural killer cells. Nature immunology *13*, 1000-1009.

Biswas, S.K., and Mantovani, A. (2010). Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nature immunology *11*, 889-896.

Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nature protocols *8*, 539-554.

Boorsma, C.E., Draijer, C., and Melgert, B.N. (2013). Macrophage heterogeneity in respiratory diseases. Mediators of inflammation *2013*, 769214.

Borchers, A.T., Keen, C.L., Huntley, A.C., and Gershwin, M.E. (2015). Lyme disease: a rigorous review of diagnostic criteria and treatment. Journal of autoimmunity *57*, 82-115.

Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S., and Lynn, D.J. (2013). InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic acids research *41*, D1228-1233.
Brightbill, H.D., Libraty, D.H., Krutzik, S.R., Yang, R.B., Belisle, J.T., Bleharski, J.R., Maitland, M., Norgard, M.V., Plevy, S.E., Smale, S.T*., et al.* (1999). Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. Science *285*, 732-736.
Burns, K., Martinon, F., Esslinger, C., Pahl, H., Schneider, P., Bodmer, J.L., Di Marco, F., French, L., and Tschopp, J. (1998). MyD88, an adapter protein involved in interleukin-1 signaling. The Journal of biological chemistry *273*, 12203-12209.
Butte, A.J., and Kohane, I.S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 418-429.
Cazzola, M., Page, C.P., Calzetta, L., and Matera, M.G. (2012). Emerging anti-inflammatory strategies for COPD. The European respiratory journal *40*, 724-741.
Cecchini, M.G., Dominguez, M.G., Mocci, S., Wetterwald, A., Felix, R., Fleisch, H., Chisholm, O., Hofstetter, W., Pollard, J.W., and Stanley, E.R. (1994). Role of colony stimulating factor-1 in the establishment and regulation of tissue macrophages during postnatal development of the mouse. Development *120*, 1357-1372.
Cervantes, J.L., Hawley, K.L., Benjamin, S.J., Weinerman, B., Luu, S.M., and Salazar, J.C. (2014). Phagosomal TLR signaling upon Borrelia burgdorferi infection. Front Cell Infect Mi *4*.
Chariot, A. (2009). The NF-kappaB-independent functions of IKK subunits in immunity and cancer. Trends in cell biology *19*, 404-413.
Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res *37*, W305-311.
Chinetti-Gbaguidi, G., and Staels, B. (2011). Macrophage polarization in metabolic disorders: functions and regulation. Current opinion in lipidology *22*, 365-372.
Choi, H., Lee, R.H., Bazhanov, N., Oh, J.Y., and Prockop, D.J. (2011). Anti-inflammatory protein TSG-6 secreted by activated MSCs attenuates zymosan-induced mouse peritonitis by decreasing TLR2/NF-kappaB signaling in resident macrophages. Blood *118*, 330-338.
Chorro, L., Sarde, A., Li, M., Woollard, K.J., Chambon, P., Malissen, B., Kissenpfennig, A., Barbaroux, J.B., Groves, R., and Geissmann, F. (2009). Langerhans cell (LC) proliferation mediates neonatal development, homeostasis, and inflammation-associated expansion of the epidermal LC network. J Exp Med *206*, 3089-3100.
Chow, A., Brown, B.D., and Merad, M. (2011). Studying the mononuclear phagocyte system in the molecular age. Nature reviews. Immunology *11*, 788-798.
Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkurst, C.N., Muratet, M*., et al.* (2012). A validated regulatory network for Th17 cell specification. Cell *151*, 289-303.
Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B*., et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nature protocols *2*, 2366-2382.
Cohen, N.R., Brennan, P.J., Shay, T., Watts, G.F., Brigl, M., Kang, J., and Brenner, M.B. (2013). Shared and distinct transcriptional programs underlie the hybrid nature of iNKT cells. Nature immunology *14*, 90-99.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.
Cortez, K.J., Lyman, C.A., Kottilil, S., Kim, H.S., Roilides, E., Yang, J., Fullmer, B., Lempicki, R., and Walsh, T.J. (2006). Functional genomics of innate host defense molecules in normal human monocytes in response to Aspergillus fumigatus. Infection and immunity *74*, 2353-2365.
Costantini, S., Capone, F., Miele, M., Guerriero, E., Napolitano, M., Colonna, G., and Castello, G. (2009). CytokineDB: a database collecting biological information. Bioinformation *4*, 92-93.
Davies, D.L., and Bouldin, D.W. (1979). A cluster separation measure. IEEE Trans Pattern Anal Mach Intell *1*, 224-227.

Davies, L.C., Jenkins, S.J., Allen, J.E., and Taylor, P.R. (2013). Tissue-resident macrophages. Nature immunology *14*, 986-995.
Davis, M.M. (2008). A prescription for human immunology. Immunity *29*, 835-838.
de Matos Simoes, R., Dehmer, M., and Emmert-Streib, F. (2013). B-cell lymphoma gene regulatory networks: biological consistency among inference methods. Frontiers in genetics *4*, 281.
De Nardo, D. (2015). Toll-like receptors: Activation, signalling and transcriptional modulation. Cytokine.
De Nardo, D., Labzin, L.I., Kono, H., Seki, R., Schmidt, S.V., Beyer, M., Xu, D., Zimmer, S., Lahrmann, C., Schildberg, F.A*., et al.* (2014). High-density lipoprotein mediates anti-inflammatory reprogramming of macrophages via the transcriptional regulator ATF3. Nature immunology *15*, 152-160.
de Souza, J.E., Galante, P.A., de Almeida, R.V., da Cunha, J.P., Ohara, D.T., Ohno-Machado, L., Old, L.J., and de Souza, S.J. (2012). SurfaceomeDB: a cancer-orientated database for genes encoding cell surface proteins. Cancer immunity *12*, 15.
Du, J., Stankiewicz, M.J., Liu, Y., Xi, Q., Schmitz, J.E., Lekstrom-Himes, J.A., and Ackerman, S.J. (2002). Novel combinatorial interactions of GATA-1, PU.1, and C/EBPepsilon isoforms regulate transcription of the gene encoding eosinophil granule major basic protein. The Journal of biological chemistry *277*, 43481-43494.
Dunay, I.R., Damatta, R.A., Fux, B., Presti, R., Greco, S., Colonna, M., and Sibley, L.D. (2008). Gr1(+) inflammatory monocytes are required for mucosal resistance to the pathogen Toxoplasma gondii. Immunity *29*, 306-317.
Dunne, A., Ejdeback, M., Ludidi, P.L., O'Neill, L.A.J., and Gay, N.J. (2003). Structural complementarity of Toll/interleukin-1 receptor domains in toll-like receptors and the adaptors Mal and MyD88. Journal of Biological Chemistry *278*, 41443-41451.
Duong, M., Ouellet, N., Simard, M., Bergeron, Y., Olivier, M., and Bergeron, M.G. (1998). Kinetic study of host defense and inflammatory response to Aspergillus fumigatus in steroid-induced immunosuppressed mice. The Journal of infectious diseases *178*, 1472-1482.
Edin, S., Wikberg, M.L., Dahlin, A.M., Rutegard, J., Oberg, A., Oldenborg, P.A., and Palmqvist, R. (2012). The distribution of macrophages with a M1 or M2 phenotype in relation to prognosis and the molecular characteristics of colorectal cancer. PloS one *7*, e47045.
Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Frontiers in cell and developmental biology *2*, 38.
Epelman, S., Lavine, K.J., Beaudin, A.E., Sojka, D.K., Carrero, J.A., Calderon, B., Brija, T., Gautier, E.L., Ivanov, S., Satpathy, A.T*., et al.* (2014a). Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. Immunity *40*, 91-104.
Epelman, S., Lavine, K.J., and Randolph, G.J. (2014b). Origin and functions of tissue macrophages. Immunity *41*, 21-35.
Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS biology *5*, e8.
Fang, F.C. (2004). Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. Nature reviews. Microbiology *2*, 820-832.
Fingerle, G., Pforte, A., Passlick, B., Blumenstein, M., Strobel, M., and Zieglerheitbrock, H.W.L. (1993). The Novel Subset of Cd14+/Cd16+ Blood Monocytes Is Expanded in Sepsis Patients. Blood *82*, 3170-3176.
Fitzgerald, K.A., Palsson-McDermott, E.M., Bowie, A.G., Jefferies, C.A., Mansell, A.S., Brady, G., Brint, E., Dunne, A., Gray, P., Harte, M.T*., et al.* (2001). Mal (MyD88-adapter-like) is required for Toll-like receptor-4 signal transduction. Nature *413*, 78-83.
Flexer, A. (1997). Limitations of self-organizing maps for vector quantization and multidimensional scaling.
Flynn, J.L., and Chan, J. (2001). Immunology of tuberculosis. Annu Rev Immunol *19*, 93-129.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using Bayesian networks to analyze expression data. J Comput Biol *7*, 601-620.
Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., and Sladek, R. (2009). TFCat: the curated catalog of mouse and human transcription factors. Genome biology *10*, R29.
Furuhashi, M., Ishimura, S., Ota, H., and Miura, T. (2011). Lipid chaperones and metabolic inflammation. International journal of inflammation *2011*, 642612.
Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S*., et al.* (2012). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nature immunology *13*, 1118-1128.
Gay, N.J., Gangloff, M., and O'Neill, L.A.J. (2011). What the Myddosome structure tells us about the initiation of innate immunity. Trends in immunology *32*, 104-109.
Giamarellos-Bourboulis, E.J., and Raftogiannis, M. (2012). The immune response to severe bacterial infections: consequences for therapy. Expert review of anti-infective therapy *10*, 369-380.
Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R*., et al.* (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science *330*, 841-845.
Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nature reviews. Immunology *14*, 392-404.
Gioannini, T.L., and Weiss, J.P. (2007). Regulation of interactions of Gram-negative bacterial endotoxins with mammalian cells. Immunologic research *39*, 249-260.
Gohda, J., Matsumura, T., and Inoue, J. (2004). Cutting edge: TNFR-associated factor (TRAF) 6 is essential for MyD88-dependent pathway but not toll/IL-1 receptor domain-containing adaptor-inducing IFN-beta (TRIF)-dependent pathway in TLR signaling. J Immunol *173*, 2913-2917.
Gordon, S., and Martinez, F.O. (2010). Alternative activation of macrophages: mechanism and functions. Immunity *32*, 593-604.
Grivennikov, S.I., Tumanov, A.V., Liepinsh, D.J., Kruglov, A.A., Marakusha, B.I., Shakhov, A.N., Murakami, T., Drutskaya, L.N., Forster, I., Clausen, B.E*., et al.* (2005). Distinct and nonredundant in vivo functions of TNF produced by t cells and macrophages/neutrophils: protective and deleterious effects. Immunity *22*, 93-104.
Guilliams, M., De Kleer, I., Henri, S., Post, S., Vanhoutte, L., De Prijck, S., Deswarte, K., Malissen, B., Hammad, H., and Lambrecht, B.N. (2013). Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. J Exp Med *210*, 1977-1992.
Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S.H., Onai, N., Schraml, B.U., Segura, E., Tussiwand, R., and Yona, S. (2014). Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. Nature reviews. Immunology *14*, 571-578.
Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., Becker, C.D., See, P., Price, J., Lucas, D*., et al.* (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. Immunity *38*, 792-804.
Hayden, M.S., and Ghosh, S. (2008). Shared principles in NF-kappaB signaling. Cell *132*, 344-362.
He, Y., Sadahiro, T., Noh, S.I., Wang, H., Todo, T., Chai, N.N., Klein, A.S., and Wu, G.D. (2009). Flow cytometric isolation and phenotypic characterization of two subsets of ED2+(CD163) hepatic macrophages in rats. Hepatol Res *39*, 1208-1218.
Heng, T.S., and Painter, M.W. (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nature immunology *9*, 1091-1094.
Hodge, S., Matthews, G., Mukaro, V., Ahern, J., Shivam, A., Hodge, G., Holmes, M., Jersmann, H., and Reynolds, P.N. (2011). Cigarette smoke-induced changes to alveolar macrophage phenotype and function are improved by treatment with procysteine. American journal of respiratory cell and molecular biology *44*, 673-681.
Hoeffel, G., Wang, Y., Greter, M., See, P., Teo, P., Malleret, B., Leboeuf, M., Low, D., Oller, G., Almeida, F*., et al.* (2012). Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. J Exp Med *209*, 1167-1181.

Hohl, T.M., Rivera, A., and Pamer, E.G. (2006). Immunity to fungi. Current opinion in immunology *18*, 465-472.
Holt, M.P., Cheng, L.L., and Ju, C. (2008). Identification and characterization of infiltrating macrophages in acetaminophen-induced liver injury. Journal of leukocyte biology *84*, 1410-1421.
Honda, K., Yanai, H., Mizutani, T., Negishi, H., Shimada, N., Suzuki, N., Ohba, Y., Takaoka, A., Yeh, W.C., and Taniguchi, T. (2004). Role of a transductional-transcriptional processor complex involving MyD88 and IRF-7 in Toll-like receptor signaling. Proceedings of the National Academy of Sciences of the United States of America *101*, 15416-15421.
Horng, T., Barton, G.M., Flavell, R.A., and Medzhitov, R. (2002). The adaptor molecule TIRAP provides signalling specificity for Toll-like receptors. Nature *420*, 329-333.
Hoshino, K., Takeuchi, O., Kawai, T., Sanjo, H., Ogawa, T., Takeda, Y., Takeda, K., and Akira, S. (1999). Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: Evidence for TLR4 as the Lps gene product. Journal of Immunology *162*, 3749-3752.
Houghton, A.M. (2013). Mechanistic links between COPD and lung cancer. Nature reviews. Cancer *13*, 233-245.
Hu, X., Chung, A.Y., Wu, I., Foldi, J., Chen, J., Ji, J.D., Tateya, T., Kang, Y.J., Han, J., Gessler, M*., et al.* (2008). Integrated regulation of Toll-like receptor responses by Notch and interferon-gamma pathways. Immunity *29*, 691-703.
Huang, J., Gong, Z., Ghosal, G., and Chen, J. (2009). SOSS complexes participate in the maintenance of genomic stability. Molecular cell *35*, 384-393.
Hume, D.A., and Freeman, T.C. (2014). Transcriptomic analysis of mononuclear phagocyte differentiation and activation. Immunological reviews *262*, 74-84.
Hume, D.A., Mabbott, N., Raza, S., and Freeman, T.C. (2013). Can DCs be distinguished from macrophages by molecular signatures? Nature immunology *14*, 187-189.
Hume, D.A., Summers, K.M., Raza, S., Baillie, J.K., and Freeman, T.C. (2010). Functional clustering and lineage markers: insights into cellular differentiation and gene function from large-scale microarray studies of purified primary cell populations. Genomics *95*, 328-338.
Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PloS one *5*.
Ibrahim-Granet, O., Philippe, B., Boleti, H., Boisvieux-Ulrich, E., Grenet, D., Stern, M., and Latge, J.P. (2003). Phagocytosis and intracellular fate of Aspergillus fumigatus conidia in alveolar macrophages. Infection and immunity *71*, 891-903.
Ingman, W.V., Glynn, D.J., and Hutchinson, M.R. (2014). Inflammatory mediators in mastitis and lactation insufficiency. Journal of mammary gland biology and neoplasia *19*, 161-167.
Jaeger, M., van der Lee, R., Cheng, S.C., Johnson, M.D., Kumar, V., Ng, A., Plantinga, T.S., Smeekens, S.P., Oosting, M., Wang, X*., et al.* (2015). The RIG-I-like helicase receptor MDA5 (IFIH1) is involved in the host defense against Candida infections. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology.
Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.
Jolliffe, I.T. (2002). Principal component analysis, 2nd edn (New York: Springer).
Kagan, J.C., Su, T., Horng, T., Chow, A., Akira, S., and Medzhitov, R. (2008). TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. Nature immunology *9*, 361-368.
Kanter, H.L., Saffitz, J.E., and Beyer, E.C. (1994). Molecular cloning of two human cardiac gap junction proteins, connexin40 and connexin45. Journal of molecular and cellular cardiology *26*, 861-868.
Karlmark, K.R., Tacke, F., and Dunay, I.R. (2012). Monocytes in health and disease - Minireview. European journal of microbiology & immunology *2*, 97-102.
Karlmark, K.R., Weiskirchen, R., Zimmermann, H.W., Gasssler, N., Ginhoux, F., Weber, C., Merad, M., Luedde, T., Trautwein, C., and Tacke, F. (2009). Hepatic Recruitment of the Inflammatory Gr1(+) Monocyte Subset Upon Liver Injury Promotes Hepatic Fibrosis. Hepatology *50*, 261-274.
Kasama, Y., Mizukami, T., Kusunoki, H., Peveling-Oberhag, J., Nishito, Y., Ozawa, M., Kohara, M., Mizuochi, T., and Tsukiyama-Kohara, K. (2014). B-cell-intrinsic hepatitis C virus expression leads to B-cell-lymphomagenesis and induction of NF-kappaB signalling. PloS one *9*, e91373.

Kierdorf, K., Erny, D., Goldmann, T., Sander, V., Schulz, C., Perdiguero, E.G., Wieghofer, P., Heinrich, A., Riemke, P., Holscher, C*., et al.* (2013). Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. Nature neuroscience *16*, 273-280.
Kim, H.S., Choi, E.H., Khan, J., Roilides, E., Francesconi, A., Kasai, M., Sein, T., Schaufele, R.L., Sakurai, K., Son, C.G*., et al.* (2005). Expression of genes encoding innate host defense molecules in normal human monocytes in response to Candida albicans. Infection and immunity *73*, 3714-3724.
Kleinnijenhuis, J., Oosting, M., Joosten, L.A., Netea, M.G., and Van Crevel, R. (2011). Innate immune recognition of Mycobacterium tuberculosis. Clinical & developmental immunology *2011*, 405310.
Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. Biol Cybern *43*, 59-69.
Kohonen, T. (2013). Essentials of the self-organizing map. Neural Networks *37*, 52-65.
Kono, H., Fujii, H., Asakawa, M., Yamamoto, M., Maki, A., Matsuda, M., Rusyn, I., and Matsumoto, Y. (2002). Functional heterogeneity of the Kupffer cell population is involved in the mechanism of gadolinium chloride in rats administered endotoxin. J Surg Res *106*, 179-187.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res *19*, 1639-1645.
Kumar, V., Abbas, A.K., Aster, J.C., and Robbins, S.L. (2013). Robbins basic pathology, 9th edn (Philadelphia, PA: Elsevier/Saunders).
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. Bmc Bioinformatics *9*.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.
Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell *159*, 1312-1326.
Lawrence, T., and Natoli, G. (2011). Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nature reviews. Immunology *11*, 750-761.
Lee, J.Y., Sohn, K.H., Rhee, S.H., and Hwang, D. (2001). Saturated fatty acids, but not unsaturated fatty acids, induce the expression of cyclooxygenase-2 mediated through Toll-like receptor 4. The Journal of biological chemistry *276*, 16683-16689.
Lin, S.C., Lo, Y.C., and Wu, H. (2010). Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. Nature *465*, 885-U882.
Mabbott, N.A., Kenneth Baillie, J., Hume, D.A., and Freeman, T.C. (2010). Meta-analysis of lineage-specific gene expression signatures in mouse leukocyte populations. Immunobiology *215*, 724-736.
Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.
Magee, D.A., Conlon, K.M., Nalpas, N.C., Browne, J.A., Pirson, C., Healy, C., McLoughlin, K.E., Chen, J., Vordermeier, H.M., Gormley, E*., et al.* (2014). Innate cytokine profiling of bovine alveolar macrophages reveals commonalities and divergence in the response to Mycobacterium bovis and Mycobacterium tuberculosis infection. Tuberculosis (Edinb) *94*, 441-450.
Maimon, O., and Rokach, L. (2010). Data mining and knowledge discovery handbook, 2nd edn (New York: Springer).
Manning, C.D., Raghavan, P., and Schütze, H. (2008). Introduction to information retrieval (New York: Cambridge University Press).
Mantovani, A., Sozzani, S., Locati, M., Allavena, P., and Sica, A. (2002). Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends in immunology *23*, 549-555.
Marakalala, M.J., Graham, L.M., and Brown, G.D. (2010). The role of Syk/CARD9-coupled C-type lectin receptors in immunity to Mycobacterium tuberculosis infections. Clinical & developmental immunology *2010*, 567571.
Marbach, D., Costello, J.C., Kuffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Kellis, M., Collins, J.J., and Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. Nature methods *9*, 796-804.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. Bmc Bioinformatics *7 Suppl 1*, S7.
Marino, M.W., Dunn, A., Grail, D., Inglese, M., Noguchi, Y., Richards, E., Jungbluth, A., Wada, H., Moore, M., Williamson, B*., et al.* (1997). Characterization of tumor necrosis factor-deficient mice. Proceedings of the National Academy of Sciences of the United States of America *94*, 8093-8098.
Marodi, L., Cypowyj, S., Toth, B., Chernyshova, L., Puel, A., and Casanova, J.L. (2012). Molecular mechanisms of mucocutaneous immunity against Candida and Staphylococcus species. The Journal of allergy and clinical immunology *130*, 1019-1027.
Martinez, F.O., Gordon, S., Locati, M., and Mantovani, A. (2006). Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. J Immunol *177*, 7303-7311.
Martinez, F.O., Helming, L., and Gordon, S. (2009). Alternative Activation of Macrophages: An Immunologic Functional Perspective. Annu Rev Immunol *27*, 451-483.
Mc Mahon, S.S., Sim, A., Filippi, S., Johnson, R., Liepe, J., Smith, D., and Stumpf, M.P. (2014). Information theory and signal transduction systems: from molecular information processing to network inference. Semin Cell Dev Biol *35*, 98-108.
McDermott, J.E., Archuleta, M., Thrall, B.D., Adkins, J.N., and Waters, K.M. (2011). Controlling the response: predictive modeling of a highly central, pathogen-targeted core response module in macrophage activation. PloS one *6*, e14673.
McGettrick, A.F., and O'Neill, L.A.J. (2010). Localisation and trafficking of Toll-like receptors: an important mode of regulation. Current opinion in immunology *22*, 20-27.
Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one *5*, e13984.
Miller, J.C., Brown, B.D., Shay, T., Gautier, E.L., Jojic, V., Cohain, A., Pandey, G., Leboeuf, M., Elpek, K.G., Helft, J*., et al.* (2012). Deciphering the transcriptional network of the dendritic cell lineage. Nature immunology *13*, 888-899.
Mitsuzawa, H., Nishitani, C., Hyakushima, N., Shimizu, T., Sano, H., Matsushima, N., Fukase, K., and Kuroki, Y. (2006). Recombinant soluble forms of extracellular TLR4 domain and MD-2 inhibit lipopolysaccharide binding on cell surface and dampen lipopolysaccharide-induced pulmonary inflammation in mice. J Immunol *177*, 8133-8139.
Miyake, K. (2007). Innate immune sensing of pathogens and danger signals by cell surface Toll-like receptors. Seminars in immunology *19*, 3-10.
Mohamed, M.R., and McFadden, G. (2009). NFkB inhibitors: strategies from poxviruses. Cell Cycle *8*, 3125-3132.
Moore, M.W., Cruz, A.R., LaVake, C.J., Marzo, A.L., Eggers, C.H., Salazar, J.C., and Radolf, J.D. (2007). Phagocytosis of Borrelia burgdorferi and Treponema pallidum potentiates innate immune activation and induces gamma interferon production. Infection and immunity *75*, 2046-2062.
Mosmann, T.R., and Coffman, R.L. (1989). TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu Rev Immunol *7*, 145-173.
Mosser, D.M., and Edwards, J.P. (2008). Exploring the full spectrum of macrophage activation. Nature reviews. Immunology *8*, 958-969.
Movita, D., Kreefft, K., Biesta, P., van Oudenaren, A., Leenen, P.J., Janssen, H.L., and Boonstra, A. (2012). Kupffer cells express a unique combination of phenotypic and functional characteristics compared with splenic and peritoneal macrophages. Journal of leukocyte biology *92*, 723-733.
Murray, P.J., and Wynn, T.A. (2011). Protective and pathogenic functions of macrophage subsets. Nature reviews. Immunology *11*, 723-737.
Myers, J.L., Well, A., and Lorch, R.F. (2010). Research design and statistical analysis, 3rd edn (New York: Routledge).
Nagai, Y., Akashi, S., Nagafuku, M., Ogata, M., Iwakura, Y., Akira, S., Kitamura, T., Kosugi, A., Kimoto, M., and Miyake, K. (2002). Essential role of MD-2 in LPS responsiveness and TLR4 distribution. Nature immunology *3*, 667-672.

Naik, S.H., Perie, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229-232.
Napetschnig, J., and Wu, H. (2013). Molecular basis of NF-kappaB signaling. Annual review of biophysics *42*, 443-468.
Nathan, C.F., Murray, H.W., Wiebe, M.E., and Rubin, B.Y. (1983). Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. J Exp Med *158*, 670-689.
Nau, G.J., Richmond, J.F., Schlesinger, A., Jennings, E.G., Lander, E.S., and Young, R.A. (2002). Human macrophage activation programs induced by bacterial pathogens. Proceedings of the National Academy of Sciences of the United States of America *99*, 1503-1508.
Netea, M.G., Wijmenga, C., and O'Neill, L.A. (2012). Genetic variation in Toll-like receptors and disease susceptibility. Nature immunology *13*, 535-542.
Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nature methods *12*, 453-457.
Nicholson, S., Bonecini-Almeida Mda, G., Lapa e Silva, J.R., Nathan, C., Xie, Q.W., Mumford, R., Weidner, J.R., Calaycay, J., Geng, J., Boechat, N.*, et al.* (1996). Inducible nitric oxide synthase in pulmonary alveolar macrophages from patients with tuberculosis. J Exp Med *183*, 2293-2302.
Ning, S., Pagano, J.S., and Barber, G.N. (2011). IRF7: activation, regulation, modification and function. Genes and immunity *12*, 399-414.
Noy, R., and Pollard, J.W. (2014). Tumor-associated macrophages: from mechanisms to therapy. Immunity *41*, 49-61.
O'Neill, L.A.J., and Bowie, A.G. (2007). The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. Nature Reviews Immunology *7*, 353-364.
Oesper, L., Merico, D., Isserlin, R., and Bader, G.D. (2011). WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. Source code for biology and medicine *6*, 7.
Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.
Ostuni, R., Kratochvill, F., Murray, P.J., and Natoli, G. (2015). Macrophages and cancer: from mechanisms to therapeutic implications. Trends in immunology *36*, 229-239.
Passlick, B., Flieger, D., and Ziegler-Heitbrock, H.W. (1989). Identification and characterization of a novel monocyte subpopulation in human peripheral blood. Blood *74*, 2527-2534.
Patel, P.S., Buras, E.D., and Balasubramanyam, A. (2013). The role of the immune system in obesity and insulin resistance. Journal of obesity *2013*, 616193.
Poland, G.A., Quill, H., and Togias, A. (2013). Understanding the human immune system in the 21st century: The Human Immunology Project Consortium. Vaccine.
Popov, A., Abdullah, Z., Wickenhauser, C., Saric, T., Driesen, J., Hanisch, F.G., Domann, E., Raven, E.L., Dehus, O., Hermann, C.*, et al.* (2006). Indoleamine 2,3-dioxygenase-expressing dendritic cells form suppurative granulomas following Listeria monocytogenes infection. The Journal of clinical investigation *116*, 3160-3170.
Popov, A., Driesen, J., Abdullah, Z., Wickenhauser, C., Beyer, M., Debey-Pascher, S., Saric, T., Kummer, S., Takikawa, O., Domann, E.*, et al.* (2008). Infection of myeloid dendritic cells with Listeria monocytogenes leads to the suppression of T cell function by multiple inhibitory mechanisms. J Immunol *181*, 4976-4988.
Qian, J., Dolled-Filhart, M., Lin, J., Yu, H.Y., and Gerstein, M. (2001a). Beyond synexpression relationships: Local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. J Mol Biol *314*, 1053-1066.
Qian, Y., Commane, M., Ninomiya-Tsuji, J., Matsumoto, K., and Li, X. (2001b). IRAK-mediated translocation of TRAF6 and TAB2 in the interleukin-1-induced activation of NFkappa B. The Journal of biological chemistry *276*, 41661-41667.
Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., Li, B., Gilchrist, M., Gold, E.S., Johnson, C.D., Litvak, V.*, et al.* (2008). Uncovering a macrophage transcriptional program by

integrating evidence from motif scanning and expression dynamics. PLoS computational biology *4*, e1000021.
Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nature genetics *38*, 500-501.
Reinartz, S., Schumann, T., Finkernagel, F., Wortmann, A., Jansen, J.M., Meissner, W., Krause, M., Schworer, A.M., Wagner, U., Muller-Brusselbach, S., and Muller, R. (2013). Mixed-polarization phenotype of ascites-associated macrophages in human ovarian carcinoma: Correlation of CD163 expression, cytokine levels and early relapse. International journal of cancer. Journal international du cancer.
Romani, L. (2004). Immunity to fungal infections. Nature reviews. Immunology *4*, 1-23.
Sagaert, X., Tousseyn, T., De Hertogh, G., and Geboes, K. (2012). Macrophage-related diseases of the gut: a pathologist's perspective. Virchows Archiv : an international journal of pathology *460*, 555-567.
Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. Nature methods *9*, 1069-1076.
Salazar, J.C., Pope, C.D., Sellati, T.J., Feder, H.M., Kiely, T.G., Dardick, K.R., Buckman, R.L., Moore, M.W., Caimano, M.J., Pope, J.G*., et al.* (2003). Coevolution of markers of innate and adaptive immunity in skin and peripheral blood of patients with erythema migrans. (vol 171, pg 2660, 2003). Journal of Immunology *171*, 4934-4934.
Sander, L.E., Davis, M.J., Boekschoten, M.V., Amsen, D., Dascher, C.C., Ryffel, B., Swanson, J.A., Muller, M., and Blander, J.M. (2011). Detection of prokaryotic mRNA signifies microbial viability and promotes immunity (vol 474, pg 385, 2011). Nature *478*, 136-136.
Sato, S., Sugiyama, M., Yamamoto, M., Watanabe, Y., Kawai, T., Takeda, K., and Akira, S. (2003). Toll/IL-1 receptor domain-containing adaptor inducing IFN-beta (TRIF) associates with TNF receptor-associated factor 6 and TANK-Binding kinase 1, and activates two distinct transcription factors, NF-kappa B and IFN-regulatory factor-3, in the toll-like receptor signaling. Journal of Immunology *171*, 4304-4310.
Satpathy, A.T., Wu, X., Albring, J.C., and Murphy, K.M. (2012). Re(de)fining the dendritic cell lineage. Nature immunology *13*, 1145-1154.
Scanga, C.A., Bafica, A., Feng, C.G., Cheever, A.W., Hieny, S., and Sher, A. (2004). MyD88-deficient mice display a profound loss in resistance to Mycobacterium tuberculosis associated with partially impaired Th1 cytokine and nitric oxide synthase 2 expression. Infection and immunity *72*, 2400-2404.
Schaffner, A., Douglas, H., Braude, A.I., and Davis, C.E. (1983). Killing of Aspergillus spores depends on the anatomical source of the macrophage. Infection and immunity *42*, 1109-1115.
Schraml, B.U., van Blijswijk, J., Zelenay, S., Whitney, P.G., Filby, A., Acton, S.E., Rogers, N.C., Moncaut, N., Carvajal, J.J., and Reis e Sousa, C. (2013). Genetic tracing via DNGR-1 expression history defines dendritic cells as a hematopoietic lineage. Cell *154*, 843-858.
Schroder, K., Irvine, K.M., Taylor, M.S., Bokil, N.J., Le Cao, K.A., Masterman, K.A., Labzin, L.I., Semple, C.A., Kapetanovic, R., Fairbairn, L*., et al.* (2012). Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages. Proceedings of the National Academy of Sciences of the United States of America *109*, E944-953.
Schultze, J.L., Freeman, T., Hume, D.A., and Latz, E. (2015). A transcriptional perspective on human macrophage biology. Seminars in immunology.
Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E., Pollard, J.W*., et al.* (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science *336*, 86-90.
Schwartz, J.T., Bandyopadhyay, S., Kobayashi, S.D., McCracken, J., Whitney, A.R., Deleo, F.R., and Allen, L.A. (2013). Francisella tularensis alters human neutrophil gene expression: insights into the molecular basis of delayed neutrophil apoptosis. Journal of innate immunity *5*, 124-136.
Serbina, N.V., and Pamer, E.G. (2006). Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. Nature immunology *7*, 311-317.
Shay, J.E.S., and Simon, M.C. (2012). Hypoxia-inducible factors: Crosstalk between inflammation and metabolism. Semin Cell Dev Biol *23*, 389-394.

Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A., and ImmGen, C. (2013). Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proceedings of the National Academy of Sciences of the United States of America *110*, 2946-2951.
Shaykhiev, R., Krause, A., Salit, J., Strulovici-Barel, Y., Harvey, B.G., O'Connor, T.P., and Crystal, R.G. (2009). Smoking-dependent reprogramming of alveolar macrophage polarization: implication for pathogenesis of chronic obstructive pulmonary disease. J Immunol *183*, 2867-2883.
Shi, C., Jia, T., Mendez-Ferrer, S., Hohl, T.M., Serbina, N.V., Lipuma, L., Leiner, I., Li, M.O., Frenette, P.S., and Pamer, E.G. (2011). Bone marrow mesenchymal stem and progenitor cells induce monocyte emigration in response to circulating toll-like receptor ligands. Immunity *34*, 590-601.
Shwetha, S., Gouthamchandra, K., Chandra, M., Ravishankar, B., Khaja, M.N., and Das, S. (2013). Circulating miRNA profile in HCV infected serum: novel insight into pathogenesis. Scientific reports *3*, 1555.
Sica, A., and Bronte, V. (2007). Altered macrophage differentiation and immune dysfunction in tumor development. The Journal of clinical investigation *117*, 1155-1166.
Sica, A., and Mantovani, A. (2012). Macrophage plasticity and polarization: in vivo veritas. The Journal of clinical investigation *122*, 787-795.
Sirnes, S., Honne, H., Ahmed, D., Danielsen, S.A., Rognum, T.O., Meling, G.I., Leithe, E., Rivedal, E., Lothe, R.A., and Lind, G.E. (2011). DNA methylation analyses of the connexin gene family reveal silencing of GJC1 (Connexin45) by promoter hypermethylation in colorectal cancer. Epigenetics : official journal of the DNA Methylation Society *6*, 602-609.
Smeekens, S.P., Ng, A., Kumar, V., Johnson, M.D., Plantinga, T.S., van Diemen, C., Arts, P., Verwiel, E.T., Gresnigt, M.S., Fransen, K*., et al.* (2013). Functional genomics identifies type I interferon pathway as central for host defense against Candida albicans. Nature communications *4*, 1342.
Smyth, G.K. (2005). limma: Linear Models for Microarray Data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, R. Gentleman, V. Carey, W. Huber, R. Irizarry, and S. Dudoit, eds. (Springer New York), pp. 397-420.
Spence, S., Fitzsimons, A., Boyd, C.R., Kessler, J., Fitzgerald, D., Elliott, J., Gabhann, J.N., Smith, S., Sica, A., Hams, E*., et al.* (2013). Suppressors of cytokine signaling 2 and 3 diametrically control macrophage polarization. Immunity *38*, 66-78.
Staal, F.J.T., van der Burg, M., Wessels, L.F.A., Barendregt, B.H., Baert, M.R.M., van den Burg, C.M.M., Van Huffel, C., Langerak, A.W., van der Velden, V.H.J., Reinders, M.J.T., and van Dongen, J.J.M. (2003). DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers. Leukemia *17*, 1324-1332.
Staschke, K.A., Dong, S.C., Saba, J., Zhao, J.Y., Brooks, N.A., Hepburn, D.L., Xia, J.Q., Gulen, M.F., Kang, Z.Z., Altuntas, C.Z*., et al.* (2009). IRAK4 Kinase Activity Is Required for Th17 Differentiation and Th17-Mediated Disease. Journal of Immunology *183*, 568-577.
Stein, M., Keshav, S., Harris, N., and Gordon, S. (1992). Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. J Exp Med *176*, 287-292.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America *102*, 15545-15550.
Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S., Honda, K., Ohba, Y., Mak, T.W., and Taniguchi, T. (2005). Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. Nature *434*, 243-249.
Takayama, K., Rothenberg, R.J., and Barbour, A.G. (1987). Absence of lipopolysaccharide in the Lyme disease spirochete, Borrelia burgdorferi. Infection and immunity *55*, 2311-2313.
Takeuchi, O., and Akira, S. (2010). Pattern recognition receptors and inflammation. Cell *140*, 805-820.
Tall, A.R., and Yvan-Charvet, L. (2015). Cholesterol, inflammation and innate immunity. Nature reviews. Immunology *15*, 104-116.

Tamoutounour, S., Guilliams, M., Montanana Sanchis, F., Liu, H., Terhorst, D., Malosse, C., Pollet, E., Ardouin, L., Luche, H., Sanchez, C*., et al.* (2013). Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. Immunity *39*, 925-938.
Tang, T., Francois, N., Glatigny, A., Agier, N., Mucchielli, M.H., Aggerbeck, L., and Delacroix, H. (2007). Expression ratio evaluation in two-colour microarray experiments is significantly improved by correcting image misalignment. Bioinformatics *23*, 2686-2691.
Theocharidis, A., van Dongen, S., Enright, A.J., and Freeman, T.C. (2009). Network visualization and analysis of gene expression data using BioLayout Express(3D). Nature protocols *4*, 1535-1550.
Thoma-Uszynski, S., Stenger, S., Takeuchi, O., Ochoa, M.T., Engele, M., Sieling, P.A., Barnes, P.F., Rollinghoff, M., Bolcskei, P.L., Wagner, M*., et al.* (2001). Induction of direct antimicrobial activity through mammalian toll-like receptors. Science *291*, 1544-1547.
Tobias, P.S., Soldau, K., and Ulevitch, R.J. (1986). Isolation of a lipopolysaccharide-binding acute phase reactant from rabbit serum. J Exp Med *164*, 777-793.
Tortora, G.J., and Derrickson, B. (2012). Principles of anatomy & physiology, 13th edn (Hoboken, NJ: Wiley).
Tranchevent, L.C., Barriot, R., Yu, S., Van Vooren, S., Van Loo, P., Coessens, B., De Moor, B., Aerts, S., and Moreau, Y. (2008). ENDEAVOUR update: a web resource for gene prioritization in multiple species. Nucleic acids research *36*, W377-384.
Tsai, M.C., Chakravarty, S., Zhu, G., Xu, J., Tanaka, K., Koch, C., Tufariello, J., Flynn, J., and Chan, J. (2006). Characterization of the tuberculous granuloma in murine and human lungs: cellular composition and relative tissue oxygen tension. Cellular microbiology *8*, 218-232.
Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T.C., Latz, E., Xiao, T.S*., et al.* (2010). IFI16 is an innate immune sensor for intracellular DNA. Nature immunology *11*, 997-U942.
van den Berg, L.M., Gringhuis, S.I., and Geijtenbeek, T.B. (2012). An evolutionary perspective on C-type lectins in infection and immunity. Annals of the New York Academy of Sciences *1253*, 149-158.
van der Maaten, L., and Hinton, G. (2012). Visualizing non-metric similarities in multiple maps. Mach Learn *87*, 33-55.
Van Dongen, S. (2008). Graph clustering via a discrete uncoupling process. Siam J Matrix Anal A *30*, 121-141.
Vanfurth, R., and Cohn, Z.A. (1968). Origin and Kinetics of Mononuclear Phagocytes. J Exp Med *128*, 415-&.
Vanfurth, R., Cohn, Z.A., Hirsch, J.G., Humphrey, J.H., Spector, W.G., and Langevoo.Hl (1972). Mononuclear Phagocytic System - New Classification of Macrophages, Monocytes and Their Strain Cells. B World Health Organ *47*, 651-658.
Verbeek, W., Gombart, A.F., Chumakov, A.M., Muller, C., Friedman, A.D., and Koeffler, H.P. (1999). C/EBPepsilon directly interacts with the DNA binding domain of c-myb and cooperatively activates transcription of myeloid promoters. Blood *93*, 3327-3337.
Vlahos, R., and Bozinovski, S. (2014). Role of alveolar macrophages in chronic obstructive pulmonary disease. Frontiers in immunology *5*, 435.
Wang, W.L., Liu, W., Gong, H.Y., Hong, J.R., Lin, C.C., and Wu, J.L. (2011). Activation of cytokine expression occurs through the TNFalpha/NF-kappaB-mediated pathway in birnavirus-infected cells. Fish & shellfish immunology *31*, 10-21.
Wen, A.Y., Sakamoto, K.M., and Miller, L.S. (2010). The role of the transcription factor CREB in immune function. J Immunol *185*, 6413-6419.
Werhli, A.V., and Husmeier, D. (2007). Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. Stat Appl Genet Mol *6*.
Wiktor-Jedrzejczak, W., and Gordon, S. (1996). Cytokine regulation of the macrophage (M phi) system studied using the colony stimulating factor-1-deficient op/op mouse. Physiological reviews *76*, 927-947.

Woodruff, P.G., Koth, L.L., Yang, Y.H., Rodriguez, M.W., Favoreto, S., Dolganov, G.M., Paquet, A.C., and Erle, D.J. (2005). A distinctive alveolar macrophage activation state induced by cigarette smoking. American journal of respiratory and critical care medicine *172*, 1383-1392.
Wright, S.D., Ramos, R.A., Tobias, P.S., Ulevitch, R.J., and Mathison, J.C. (1990). CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. Science *249*, 1431-1433.
Wright, S.D., Tobias, P.S., Ulevitch, R.J., and Ramos, R.A. (1989). Lipopolysaccharide (LPS) binding protein opsonizes LPS-bearing particles for recognition by a novel receptor on macrophages. J Exp Med *170*, 1231-1241.
Wu, J.J. (2012). Advances in k-means clustering (New York: Springer).
Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. Nature *496*, 445-455.
Xue, J., Schmidt, S.V., Sander, J., Draffehn, A., Krebs, W., Quester, I., De Nardo, D., Gohel, T.D., Emde, M., Schmidleithner, L*., et al.* (2014). Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity *40*, 274-288.
Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., and Akira, S. (2003a). Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. Science *301*, 640-643.
Yamamoto, M., Sato, S., Hemmi, H., Uematsu, S., Hoshino, K., Kaisho, T., Takeuchi, O., Takeda, K., and Akira, S. (2003b). TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway. Nature immunology *4*, 1144-1150.
Yang, X., Chen, Y., and Gabuzda, D. (1999). ERK MAP kinase links cytokine signals to activation of latent HIV-1 infection by stimulating a cooperative interaction of AP-1 and NF-kappaB. The Journal of biological chemistry *274*, 27981-27988.
Yao, C.J., Works, K., Romagnoli, P.A., and Austin, G.E. (2005). Effects of overexpression of HBP1 upon growth and differentiation of leukemic myeloid cells. Leukemia *19*, 1958-1968.
Yeung, M.K.S., Tegner, J., and Collins, J.J. (2002). Reverse engineering gene networks using singular value decomposition and robust regression. Proceedings of the National Academy of Sciences of the United States of America *99*, 6163-6168.
Yona, S., Kim, K.W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A*., et al.* (2013). Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. Immunity *38*, 79-91.
Yosef, N., Shalek, A.K., Gaublomme, J.T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgolli, M*., et al.* (2013). Dynamic regulatory network controlling TH17 cell differentiation. Nature *496*, 461-468.
Zhang, H., and Singer, B. (2010). Recursive partitioning and applications, 2nd edn (New York: Springer).
Ziegler-Heitbrock, L. (2014). Monocyte subsets in man and other species. Cellular immunology *289*, 135-139.
Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D.N., Leenen, P.J., Liu, Y.J., MacPherson, G., Randolph, G.J*., et al.* (2010). Nomenclature of monocytes and dendritic cells in blood. Blood *116*, e74-80.
Zoppoli, P., Morganella, S., and Ceccarelli, M. (2010). TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. Bmc Bioinformatics *11*, 154.

Appendix

A. Scripts

Script 1 Implementation of Borda Count method

```
#!/usr/bin/perl
# The script aims to combine rankings from two methods by Borda Count method.

use strict;
use warnings;
my @matrix1 = ();
my @matrix2 = ();
my ($total, $fn_in1, $weight1, $fn_in2, $weight2, $fn_out) = @ARGV;

my $m = 0;
open(IN1, $fn_in1) or die "$!\n";
while (<IN1>) {
   chomp;
   my @line = ($_,$total-$m,$weight1);
   push(@matrix1,\@line);
   $m++;
}
close IN1;

my $n = 0;
open(IN2, $fn_in2) or die "$!\n";
while (<IN2>) {
   chomp;
   my @line = ($_,$total-$n,$weight2);
   push(@matrix2,\@line);
   $n++;
}
close IN2;

my %hash = ();
my @cell = (\@matrix1,\@matrix2);
for (my $i = 0;$i < @cell;$i++){
   foreach my $e (@{$cell[$i]}){
            my $name = ${$e}[0];
            my $score = ${$e}[2]*${$e}[1];
      for (my $j = 0;$j<@cell;$j++){
               if($j != $i){
                        for(my $f=0; $f < @{$cell[$j]};$f++){
                     if ($name eq ${${$cell[$j]}[$f]}[0]){
                              $score += ${${$cell[$j]}[$f]}[2] * ${${$cell[$j]}[$f]}[1];
                              splice (@{$cell[$j]}, $f,1);  #delete repetitive hits from other samples
                              last;
                      }
                     elsif ( $name =~ /(${${$cell[$j]}[$f]}[0])/){
                              ${$e}[0] = $1;
                              $score += ${${$cell[$j]}[$f]}[2] * ${${$cell[$j]}[$f]}[1];
                              splice (@{$cell[$j]}, $f,1);  #delete repetitive hits from other samples
                              last;
                             }
                  }
               }
```

```
        }
        $hash{$name} = $score;
    }
}
open(OUT, ">".$fn_out) or die "$!\n";
print OUT "Gene Symbol\tScore\n";
foreach my $key (sort hashValueDescendingNum (keys(%hash))) {
            print OUT "$key\t$hash{$key}\n";
}
close OUT;

sub hashValueDescendingNum {
   $hash{$b} <=> $hash{$a};
}
```

Script 2 Generation of input files for Circos

```
#!/usr/bin/perl
#The script aims to generate specific a batch of input files for Circos.

use strict;
use warnings;

my ($n, $m, $boolean_text, $fn_in, $fn_karyotype, $fn_link, $fn_text,$fn_hist, $fn_band) = @ARGV;
my @a = ();
my @conditions=();
my @link = ();
my @text = ();
my @hist = ();
my @count = (0)x $n;
my $intermediate = 0;
my $all = 0;

open (IN, $fn_in) or die "$!\n";
while (<IN>) {
   chomp;
   my @temp = split("\t", $_);
   push(@a, \@temp);
}
close IN;

for(my $i = 2;$i<@{$a[0]};$i++){
        push(@conditions, $a[0][$i+1]);
}

for(my $i = 1;$i<@a;$i++){
        my $present = 0;
        for(my $j = 3;$j<@{$a[$i]}-1; $j++){
          if ($a[$i][$j]>0){
                $present++;
                my $chr = $j-2;
                my $start = $count[$chr-1]++;
                my @line1 = ('hs0',$a[$i][1],$a[$i][2],'hs'."$chr",$start,$start+1);
                push (@link,\@line1);
                if ($boolean_text){
            my @line2 = ('hs'."$chr",$start,$start+1,$a[$i][0]);
            push (@text,\@line2);
    }
          }
        }
        if ($present != $a[$i][-1]){
                print "wrong count!\n"
                }
        if ($present >=1 && $present < $n){
                $intermediate++;
                }
        elsif ($present == $n){
                $all++;
                }
        my @line3 = ('hs0',$a[$i][1],$a[$i][2],$present);
   push (@hist,\@line3);

}
```

```
my $none =  $m-$intermediate-$all;
if ($none>0){
  for (my $i = 0; $i < $none;$i++){
                        my @line = ('hs0',$i,$i+1,0);
                        push(@hist,\@line);
 }
}

open(OUT1, ">".$fn_karyotype) or die "$!\n";

print OUT1 "chr\t-\ths0\tpresent\t0\t$m\tlblue\n";
if ($none>0){
            print OUT1 "band\ths0\tnon-increased\tnon-increased\t0\t$none\tblack\n";
            my $temp=$none+$intermediate;
            print OUT1 "band\ths0\tintermediate\tintermediate\t$none\t$temp\tvlgrey\n";
   print OUT1 "band\ths0\tall-increased\tall-increased\t$temp\t$m\tred\n";
}
else{
            print OUT1 "band\ths0\tintermediate\tintermediate\t0\t$intermediate\tvlgrey\n";
   print OUT1 "band\ths0\tall-increased\tall-increased\t$intermediate\t$m\tred\n";
            }
print OUT1 "chr\t-\ths1\t$conditions[0]\t0\t$count[0]\tgrey1\n";
print OUT1 "chr\t-\ths2\t$conditions[1]\t0\t$count[1]\tpink2\n";
print OUT1 "chr\t-\ths3\t$conditions[2]\t0\t$count[2]\tred3\n";
print OUT1 "chr\t-\ths4\t$conditions[3]\t0\t$count[3]\tyellow4\n";
print OUT1 "chr\t-\ths5\t$conditions[4]\t0\t$count[4]\torange5\n";
print OUT1 "chr\t-\ths6\t$conditions[5]\t0\t$count[5]\tgreen6\n";
print OUT1 "chr\t-\ths7\t$conditions[6]\t0\t$count[6]\tblue7\n";
print OUT1 "chr\t-\ths8\t$conditions[7]\t0\t$count[7]\tpurple8\n";
close OUT1;

print "Band counts\nnon-increased:$none\tintermediate:$intermediate\tall-increased:$all\n";
print "chromosome sizes:\n";
foreach my $e (@count){
            print "$e\t";
            }
print "\n";

open(OUT2, ">".$fn_link) or die "$!\n";
for (my $i = 0; $i < @link; $i++){
 for (my $j = 0; $j < @{$link[$i]}; $j++){
   print OUT2 "$link[$i][$j]\t";
 }
 print OUT2 "\n";
}
close OUT2;

if ($boolean_text){
  open(OUT3, ">".$fn_text) or die "$!\n";
  for (my $i = 0; $i < @text; $i++){
    for (my $j = 0; $j < @{$text[$i]}; $j++){
     print OUT3 "$text[$i][$j]\t";
    }
    print OUT3 "\n";
  }
  close OUT3;
}
```

```
open(OUT4, ">".$fn_hist) or die "$!\n";
for (my $i = 0; $i < @hist; $i++){
  for (my $j = 0; $j < @{$hist[$i]}; $j++){
    print OUT4 "$hist[$i][$j]\t";
  }
  print OUT4 "\n";
}
close OUT4;

open(OUT5, ">".$fn_band) or die "$!\n";
print OUT5 "hs0\t0\t$none\tnon-increased\n";
print OUT5 "hs0\t$none\t$temp\tintermediate\n";
print OUT5 "hs0\t$temp\t$m\tall-increased\n";
close OUT5;
```

Script 3 Configuration of Circos run

```
#This configuration file was used to run Circos for the Surfaceome plot (Figure 6.4.8)
karyotype =./karyotype-surfaceome.txt

chromosomes               = hs0;hs1;hs2;hs3;hs4;hs5;hs6;hs7;hs8
chromosomes_order_by_karyotype = yes
chromosomes_units         = 1000
chromosomes_display_default = yes
chromosomes_reverse= hs1;hs2;hs3;hs4;hs5;hs6;hs7;hs8

<plots>
<plot>
type = text
file = ./band-surfaceome.txt
color = black
label_font=default
label_size=40
r0   = 1.02r
r1   = 1.3r
label_parallel = yes
</plot>

<plot>
type     = histogram
file     = ./histogram-surfaceome.txt

r1       = 0.98r
r0       = 0.92r
# extend_bin  = no

<rules>
<rule>
condition = var(value) < 8
color     = black
fill_color = grey
</rule>

<rule>
condition  = var(value) > 7
color = red
fill_color = lred
</rule>
</rules>

</plot>

<plot>
type           = text
color          = black
file           = ./text-surfaceome.txt
r0 = 0.92r
r1 = 0.99r
label_size   = 20p
label_font   = condensed
padding  = 0p
rpadding = 0p
</plot>
```

```
</plots>

<colors>
# This is included from the Circos distribution. You do not need to create this file.
<<include ./colors.conf>>
# If you are using v0.55 or earlier you'll need the next line
# <<include etc/brewer.conf>>
</colors>
<fonts>
# This is included from the Circos distribution.
<<include /home/marc/circos-0.62/etc/fonts.conf>>
</fonts>
<patterns>
# This is included from the Circos distribution.
# Only useful if you are using patterned fills.
<<include /home/marc/circos-0.62/etc/patterns.conf>>
</patterns>

<<include /home/marc/circos-0.62/etc/ideogram.conf>>
<<include /home/marc/circos-0.62/etc/ticks.conf>>
<<include /home/marc/circos-0.62/etc/housekeeping.conf>>

<image>
dir   = .
file  = surfaceome
png   = yes
svg   = yes
# radius of inscribed circle in image
radius        = 2000p
#background     white
# by default angle=0 is at 3 o'clock position
angle_offset   = -135
#angle_orientation = counterclockwise
auto_alpha_colors = yes
auto_alpha_steps  = 5
</image>

<links>
show         = yes
<link>
file        = ./link-surfaceome.txt
z           = 0
radius      = 0.90r
bezier_radius = 0r
thickness    = 1p
bezier_radius_purity = 0.5
crest              = 0.25
<rules>
flow = continue
<rule>
condition    = from(hs0)
radius1      = 0.89r
</rule>
<rule>
condition    = to(hs0)
radius2      = 0.89r
</rule>
<rule>
```

```
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs1")
color      = grey1
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs2")
color      = pink2
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs3")
color      = red3
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs4")
color      = yellow4
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs5")
color      = orange5
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs6")
color      = green6
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs7")
color      = blue7
</rule>
<rule>
condition    = ( _CHR1_ eq "hs0" && _CHR2_ eq "hs8")
color      = purple8
</rule>
</rules>
</link>
</links>
```

B. Tables

Table 1: Identification of publications associated with the major hub genes using pubatlas.org

	Hub genes	FABP5	TFNAIP6	CXCR7	NME1	ACOT7	DCTPP1	MRPL24	RRP15	ATIC	GPD1L
Cell types / organs	**Total**	470	192	252	918	8	1	1	2	182	10
Macrophage	232444	25	18	8	6	1					
Dendritic cell	79457	6	4	4							
Myeloid cell	299999	23	19	14	24	1				3	
Mast cell	38269		1	2							
Kupffer cell	7778	1									
Microglia	16077	1		2							
Alveolar macrophage	16836	3		1							
Stem cell	220821	39	14	42	11					1	
T cell	299999	6	5	26	14					5	
B cell	151858	3	1	15	18					8	
NK cell	50732		1	1	4						
Granulocyte	156204		5	7	7					1	
Neutrophil	117660	1	11	3						1	
Basophil	10338										
Eosinophil	36599			1							
Heart	299999	8	3	6	2	1					5
Lung	299999	9	8	20	69					4	1
Liver	299999	54	5	13	40	1			1	4	
Brain	299999	16		32	5	1				2	
Kidney	299999	8	8	17	22		1			11	
Muscle	299999	20	8	6	11	2				4	3
Fat tissue	98308	109	1		2	1					

Numbers reflect publications with the respective search terms in columns and rows. Table entries show cumulative publications through 2012 (as accessed Mon Apr 1 07:26:39 2013 GMT).

Table 2: TFs identified among the top 10% hub genes, their predicted binding sites at gene loci of the Top 10% hub genes

TF	Degree	TF family	overrepresented z-score for TFBS	p-value from normal distribution
JUNB	33	V$AP1F	-2.88	0.003976752
BATF3	35	V$AP1F	-2.88	0.003976752
NFE2	31	V$AP1R	-1.66	0.096914453
CEBPD	33	V$CEBP	-4.9	9.58E-07
CREB1	33	V$CREB	3.58	0.000343594
E2F3	33	V$E2FF	29.3	1.04E-188
MXD4	77	V$EBOX	11.11	1.12E-28
ETS2	45	V$ETSF	7.85	4.16E-15
RIT1	36	V$EVI1	-10.23	1.46E-24
CHES1	35	V$FKHD	-12.52	5.80E-36
FOXO3	61	V$FKHD	-12.52	5.80E-36
HEY1	37	V$HESF	10.71	9.14E-27
HIF1A	50	V$HIFF	8.74	2.33E-18
MUM1	31	V$IRFF	-2.34	0.01928374
MTF1	34	V$MTF1	5.14	2.75E-07
NME1	114	V$NDPK	14.11	3.30E-45
NFKB1	30	V$NFKB	7.4	1.36E-13
HIVEP1	31	V$NFKB	7.4	1.36E-13
SMAD4	31	V$SMAD	4.64	3.48E-06
HBP1	33	V$SORY	-15.3	7.65E-53
HMGA1	43	V$SORY	-15.3	7.65E-53
CKLF	32	V$SP1F	25.86	1.88E-147
STAT4	57	V$STAT	-2.33	0.019806151
STAT3	37	V$STAT	-2.33	0.019806151
NFX1	32	V$XBBF	3.28	0.001038071
ZNF281	40	V$ZF02	24.29	2.50E-130
ZNF148	35	V$ZF02	24.29	2.50E-130