From the Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden

Experimental and computational analysis of human GM-CSF producing T_{helper} cells

Szabolcs Éliás

Karolinska Institutet

Stockholm 2018

Published by Karolinska Institutet.
Printed by Eprint AB 2017/18

ISBN 978-91-7676-921-8

Experimental and computational analysis of human GM-CSF producing T_{helper} cells

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Szabolcs Éliás

Principal Supervisor:
Professor Jesper Tegnér
Karolinska Institutet
Department of Medicine, Solna
Unit of Computational Medicine

Co-supervisor(s):
Associate Professor John Andersson
Karolinska Institutet
Department of Medicine, Solna
Unit of Immunology and Allergy

Opponent:
Associate Professor Chris Cotsapas
Yale School of Medicine
Department of Neurology & Department of Genetics

Examination Board:
Associate Professor Carsten Daub
Karolinska Institutet
Department of Biosciences and Nutrition
Clinical Transcriptomics Research Group

Professor Ola Winqvist
Karolinska Institutet
Department of Medicine, Solna
Unit of Immunology and Allergy

Professor Mikael Sigvardsson
Linköping University
Department of Clinical and Experimental Medicine;
Lund University
Faculty of Medicine, Department of Laboratory Medicine

“Keep in mind that imagination is at the heart of all innovation. Crush or constrain it and the fun will vanish.”

Albert-László Barabási

ABSTRACT

T_{helper} cells are crucial elements of the immune system, and they differentiate into several subsets depending on their cytokine environment. Each subset contributes to a certain type of immune response by producing its characteristic cytokines. Classification of T_{helper} cells into subsets is a useful conceptual framework to investigate their biological functions. However, this classification is a simplification because the subsets are rather a continuum than discrete types. A necessary condition for health is the balanced presence and activity of the different T_{helper} subsets. Imbalanced activity of the T_{helper} subsets contributes to several diseases ranging from cancer to autoimmune and inflammatory diseases.

This thesis focuses on studying T_{helper} subsets that have been previously described as being associated with and contributing to autoimmune diseases, for example Multiple Sclerosis. T_{helper} cells producing the cytokines GM-CSF, IFN-γ and/or IL-17 have been described to be important in the pathogenesis of this disease. The biological aspects studied in this thesis include the differentiation, cytokine profiles and gene regulatory patterns of these T_{helper} subsets. From a methodological point of view, this thesis also explores possibilities to combine experimental and computational approaches.

Paper I is focused on the (*in vitro*) differentiation of human GM-CSF producing T_{helper} cells. Herein, various types of stimuli are tested and analyzed in a data driven way. As a main result, the cytokine TGF-β is identified as a context dependent modulatory factor that can induce or repress the differentiation of human GM-CSF producing T_{helper} cells depending on activation type or sodium chloride concentration. GM-CSF production is highly correlated with IFN-γ on the single cell level and with FOXP3 on the population level. Furthermore, human GM-CSF producing T_{helper} cells comprise several subpopulations, the composition of which is altered by the cytokine environment.

Paper II explores the role of splice isoforms of FOXP3 (the key transcription factor of immunosuppressive regulatory T cells) in Crohn's disease and in the differentiation of human inflammatory T_{helper} cell subsets. This paper identifies a connection between the pro-inflammatory cytokine IL-1β, FOXP3 alternative splicing, and the differentiation of pro-inflammatory IL-17 producing T_{helper} cells. Furthermore, the paper reveals a significant correlation between a certain FOXP3 splice isoform and IL-17 expression in affected tissue from Crohn's disease patients.

Paper III presents a web application that allows non-expert users to apply pre-processing and advanced statistical methods on single cell cytometry data. The aim of this tool is to

make the statistical approach to cytometry data more accessible to wet-lab biologists, and therefore serve this unmet need.

Paper IV aims to give an insight and understanding into the gene regulation and the connection between chromatin and transcriptional activity in human T_{helper} cells. Herein, the focus is on studying the characteristics of memory and GM-CSF producing T_{helper} cells. For this purpose, chromatin (ATAC-seq) and gene expression (RNA-seq) data are utilized in a combined manner and gene regulatory networks, including transcription factors that appear to be important for defining memory and GM-CSF producing T_{helper} cells, are identified. These transcription factors might be involved in diseases such as Multiple Sclerosis, and they can be potential candidates for future research towards therapeutic goals.

In summary, the thesis aims to contribute to our understanding of human T cell biology that is relevant for disease by combining experimental cellular immunology, next generation sequencing and computational approaches.

LIST OF SCIENTIFIC PAPERS

I. **Éliás S**, Schmidt A, Kannan V, Andersson J, Tegnér J.
TGF-β Affects the Differentiation of Human GM-CSF+ CD4+ T Cells in an Activation- and Sodium-Dependent Manner.
Frontiers in Immunology (2016) **7**:603. doi:10.3389/fimmu.2016.00603

II. Mailer RKW, Joly A-L, Liu S, **Éliás S**, Tegnér J, Andersson J.
IL-1β promotes Th17 differentiation by inducing alternative splicing of FOXP3.
Scientific Reports (2015) **5**:14674. doi:10.1038/srep14674

III. Papoutsoglou G, Athineou G, Lagani V, Xanthopoulos I, Schmidt A, **Éliás S**, Tegnér J, Tsamardinos I.
SCENERY: a web application for (causal) network reconstruction from cytometry data.
Nucleic Acids Research (2017) **45**:W270-W275. doi:10.1093/nar/gkx448

IV. **Éliás S**, Schmidt A, Gomez-Cabrero D, Tegnér J.
Gene regulatory network of human naïve and memory T helper cells focused on GM-CSF producing cells.
Manuscript

ADDITIONAL PUBLICATIONS

Publications during doctoral studies that are not included the thesis

Schmidt A, Marabita F, Kiani NA, Gross CC, Johansson H, **Éliás S**, Rautio S, Eriksson M, Fernandez SJ, Silberberg G, Ullah U, Bhatia U, Lähdesmäki H, Lehtiö J, Gomez-Cabrero D, Wiendl H, Lahesmaa R, Tegnér J.
Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3.
Genome Biology, in revision

Schmidt A, Rieger CC, Venigalla RK, **Éliás S**, Max R, Lorenz H-M, Gröne H-J, Krammer PH, Kuhn A.
Analysis of FOXP3+ regulatory T cell subpopulations in peripheral blood and tissue of patients with systemic lupus erythematosus.
Immunologic Research (2017) **65**:551–563. doi:10.1007/s12026-017-8904-4

Schmidt A, **Éliás S**, Joshi RN, Tegnér J.
In Vitro Differentiation of Human CD4+FOXP3+ Induced Regulatory T Cells (iTregs) from Naïve CD4+ T Cells Using a TGF-β-containing Protocol.
Journal of Visualized Experiments (2016) doi:10.3791/55015

CONTENTS

1 Introduction 1
1.1 T cells in the immune system 1
1.2 T_{helper} cell subsets 4
1.3 T_{helper} cells in Multiple Sclerosis (MS) 6
1.3.1 MS: an autoimmune disease 6
1.3.2 Characterization and differentiation of pathogenic auto-reactive T_{helper} cells 7
1.3.3 GM-CSF as a crucial cytokine in MS 9
1.3.4 Induction and phenotype of GM-CSF producing CD4+ T cells 11
2 Aims 13
3 Results 15
3.1 Computational flow cytometry studies on the differentiation of pro-inflammatory T_{helper} subsets 15
3.1.1 Factors affecting GM-CSF producing T_{helper} cell differentiation – Paper I 15
3.1.2 Positioning GM-CSF producing T_{helper} cells in the T_{helper} space – Paper I 18
3.1.3 Regulation & function of FOXP3 isoforms in T_{helper} cells – Paper II 20
3.1.4 A web based tool for computational analysis of single cell cytometry data – Paper III 22
3.2 Molecular profile & gene regulatory network of *ex vivo* memory & GM-CSF producing T_{helper} cells – Paper IV 23
3.2.1 Transcriptome analysis: differential expression & disease enrichment 24
3.2.2 Prediction of transcription factor binding 25
3.2.3 Data integration: directed network reconstruction from chromatin & transcriptome information 25
4 Conclusions 29
5 Acknowledgements 31
6 References 33

LIST OF ABBREVIATIONS

APC	antigen-presenting cell
ATAC-seq	assay for transposase-accessible chromatin using sequencing
CD	cluster of differentiation
CNS	central nervous system
CNV	copy number variation
CSF2	colony stimulating factor 2; gene encoding for GM-CSF
CTL	cytotoxic T lymphocyte
EAE	experimental autoimmune encephalomyelitis
FOXP3	forkhead box P3
GM-CSF	granulocyte-macrophage colony-stimulating factor
GSEA	gene set enrichment analysis
IFN	interferon
IL	interleukin
LASSO	least absolute shrinkage and selection operator
MAIT	mucosal associated invariant T cell
MHC	major histocompatibility complex
MS	multiple sclerosis
NKT	natural killer T cell
RA	rheumatoid arthritis
RNA-seq	RNA sequencing
SNP	single-nucleotide polymorphism
STAT	signal transducer and activator of transcription
TCR	T cell receptor
TGF	transforming growth factor
Th	T_{helper} cell
TNF	tumor necrosis factor
Treg	regulatory T cell

1 INTRODUCTION

1.1 T cells in the immune system

The immune system's main function is to protect the body against foreign structures by eliminating those without harming the body's own structures (referred to as "self"). The immune system has two major arms: the innate and the adaptive immune system. The innate immune system recognizes general patterns of foreign structures, for example common molecules in bacterial cell walls, and responds rapidly to it [reviewed in (Janeway and Medzhitov, 2002)]. In contrast, the adaptive immune system (also called acquired immune system) has the ability to "adapt" to recognize virtually any new and unique structures referred to as antigens, rendering it highly specific for a certain pathogen. Due to this initial adaptation to new foreign structures, the adaptive immune system responds slow on the first encounter however, it has the ability to form immunological memory which ensures a fast and more efficient response on the second encounter with the same antigen.

Lymphocytes, a type of white blood cells (leukocytes), contribute to both innate immunity and adaptive immunity. T and B lymphocytes (also referred to as T and B cells) are the main cell types mediating the adaptive immune response, either in a cell-mediated way (T cells) or by conferring humoral immunity through secretion of antibody proteins (B cells) respectively.

The most abundant types of T lymphocytes and those involved in adaptive immunity are T_{helper} (Th) cells and cytotoxic T lymphocytes (CTLs), defined by the expression of the cluster of differentiation (CD) molecules, CD4 and CD8 respectively. The main function of T_{helper} cells is to co-ordinate immune responses by "helping" other immune cells, while CTLs can directly kill target cells such as virus-infected cells. Additional T lymphocyte lineages have been identified that display features of both innate and adaptive immunity; these include natural killer T (NKT) cells, mucosal associated invariant T (MAIT) cells and gamma delta T cells (γδ T cells), and are not further discussed here.

Almost every individual Th cell and CTL expresses a unique T cell receptor (TCR) comprised of a TCRα, a TCRβ, and CD3 chains, which enables them to recognize a specific antigen. This large repertoire of unique TCRs is generated during the development of T cells in the thymus, in which segments of TCRα and TCRβ chain genes are rearranged and those T cells with a functional rearranged TCR are positively selected, unless their affinity to self-antigens is too high which results in deletion (negative selection) of these potentially autoreactive T cells. The antigen recognized by a TCRα/β T cell is a peptide

derived from a protein of a foreign or self structure. The antigenic peptide has to be presented to the T cell by other cells, on the surface of which, the peptide is bound to a molecular complex called major histocompatibility complex (MHC), and the peptide:MCH complex is then recognized by the cognate TCR. There are two main types of MHC molecules: MHC class I and MHC class II, and the peptides bound by them are recognized by Th cells and CTLs respectively, through binding of the CD8 and CD4 co-receptors to MHC class I and class II respectively. Since MHC class I is expressed abundantly on virtually any cell type, CD8+ T cells can directly kill any infected cell with its MHC class I containing the cognate antigen. In contrast, MHC class II is expressed on so-called antigen presenting cells (APCs), including dendritic cells, macrophages and B cells, thus in order for CD4+ T cells to recognize the antigen, it needs to be presented to the Th cell by an APC, bound to the MHC class II complex.

Antigen recognition provides the T cell (both CD4+ and CD8+ T cell) with a signal; however, this is not sufficient to activate the T cell – which is important in order to limit self-reactivity to frequently presented self-antigens. Given that a dangerous situation such as an infection is associated with the antigen, additional signals are provided to the T cell: co-stimulatory molecules of the B7 family that signal to the CD28 molecule on T cells, as well as cytokines (Figure 1). Provided that these signals are present, together with the cognate antigen, the reactive T cell will enter to a process of clonal expansion and differentiation. At this point, naïve T_{helper} cells can differentiate into several subsets that fulfill different functions (see 1.2 below). Once a clone of T cells carried out their tasks (killing for CTLs, and cytokine production for T_{helper} cells) and the infection is cleared, most of them die which reduces the size of the clonal T cell population; however, some of them develop into long-lived memory T_{helper} cells. The next time these memory T cells recognize their antigen and become activated, they will react faster and act as effector memory T cells (Figure 1).

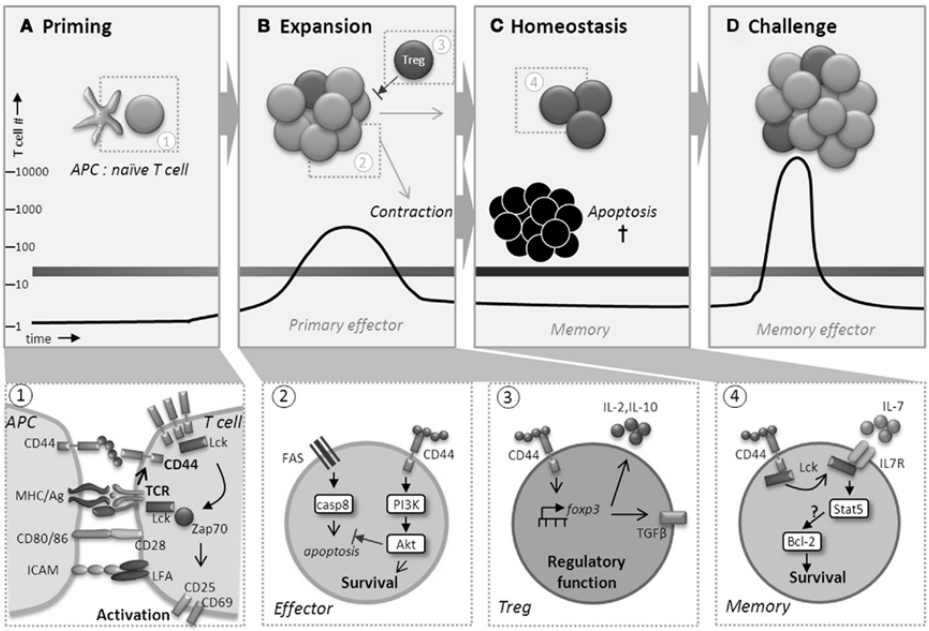

Figure 1: Schematic overview of T cell activation and memory formation.

The phases of T cell activation and memory formation are shown, with the changes in cell population size, cell-to-cell interactions and important signaling events indicated. Modified from (Baaten et al., 2012) according to the terms of the Creative Commons Attribution Non Commercial License (CC BY-NC 3.0; https://creativecommons.org/licenses/by-nc/3.0/).

1.2 T_{helper} cell subsets

Broadly, CD4+ T cells can be classified into "pro-inflammatory" effector T_{helper} cells contributing to an immune response, or "anti-inflammatory" regulatory T cells (Tregs) suppressing the immune response by inhibiting conventional T cells and multiple other immune cell types.

The major Treg subset consists CD4+ T cells expressing the "lineage-defining" transcription factor Forkhead box P3 (FOXP3), which again can be divided into thymus-derived and peripherally induced Tregs [reviewed in (Benoist and Mathis, 2012)]. Thymic-derived Tregs mature in the thymus, while induced Tregs can differentiate from naïve T cells given certain signals in the periphery (Figure 2). Besides central tolerance mentioned above that ensures negative selection of autoreactive T cells during thymic development, Tregs form an important part of peripheral tolerance that helps controlling autoreactive T cells that escaped negative selection. While this is important to prevent autoimmune diseases, on the other hand, Tregs can also contribute to prevention of anti-tumor immune responses. In the following, T_{helper} cells will refer to effector T_{helper} cells.

The main function of T_{helper} cells is to co-ordinate immune responses. They do this mostly by sending signals to other immune cells (e.g. B cells, CD8+ T cells, macrophages and neutrophils) in the form of secreted cytokines. Cytokines are proteins that can be produced by many different cell types, and which act on cells expressing the suitable receptor in an autocrine, paracrine or endocrine fashion. Most cytokine receptors signal through the JAK-STAT pathway, with certain members of the signal transducer and activator of transcription (STAT) family preferentially mediating the response to certain cytokines [reviewed in (Levy and Darnell, 2002)]. The specific combination of cytokines that a T_{helper} cell produces is also referred to as cytokine profile. The cytokine profile of a T_{helper} cell will depend on the "instructions" (cytokines and other signals) it receives from the surroundings, majorly from dendritic cells and other immune cells. These signals, in turn, depend on the type of invader these cells sensed, i.e. what the immune system has to fight (Figure 2). Consequently, there are different types of T_{helper} cells, generated for a given defense task, and they are defined by their cytokine profile, fulfilling different functions. For the major "classical" T_{helper} subsets, there is also a corresponding "lineage-defining" transcription factor that regulates the establishment of the given phenotype and cytokine profile (Figure 2). A crucial factor for the healthy immune system is a balance between the different types of T_{helper} cells. In case a given type is over-represented or over-active, it can lead to diseases.

The classical subsets of T_{helper} cells are Th1, Th2, Th17 and Tfh cells (Figure 2). Th1 cells take part in defense against intracellular viruses and bacteria, produce the signature cytokine interferon-γ (IFN-γ) along with tumor necrosis factor (TNF)-α and interleukin-2 (IL-2), and express the lineage-defining transcription factor T-bet. Th2 cells take part in defense against extracellular parasites and they produce IL-4 (as well as IL-5 and IL-13), and they express GATA3. Th17 cells take part in defense against extracellular pathogens, such as fungi and bacteria, and they produce IL-17 (as well as IL-21) and express RORγ(t) [reviewed in (Romagnani, 2014)]. Tfh (T follicular helper) cells provide help to B cells for antibody production and they are mainly located in the follicles. They produce IL-21 but they can also produce Th1 and Th2 cytokines, and they express the lineage-defining transcription factor Bcl6 [reviewed in (Baumjohann and Ansel, 2014; Crotty, 2014)].

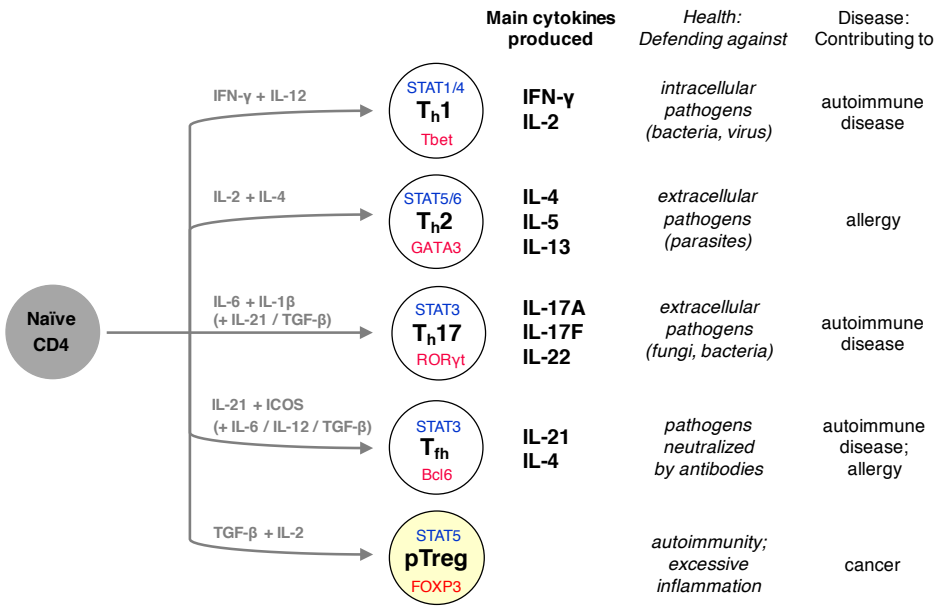

Figure 2: A simplified view of the major subsets of CD4+ T_{helper} cells.
Naïve CD4+ T cells (depicted in grey) can differentiate into different subsets, when stimulated via the TCR in the presence of co-stimulation. The major subsets are the pro-inflammatory effector subsets Th1, Th2, Th17, Tfh (white) as well as anti-inflammatory Tregs (yellow). Inducing cytokines or other signals are given in grey. Inducing cytokines in parentheses were found to be necessary only in some of the studies, and/or differ between mouse and human. The "master transcription factor" for each subset is given in red, and the major STAT family members downstream of cytokine signaling receptors in blue. The major cytokines produced by each subset are indicated in black. In the healthy state, each effector subset contributes to the defense of certain types of pathogens, while Tregs inhibit the immune response and hence prevent autoimmune disease and excessive inflammation. Over-activity of the different subsets, on the contrary, can contribute to diseases such as allergy and autoimmune disease, as indicated.

Dividing T_{helper} cells into subsets is certainly a useful model when thinking about their biological roles and functions, but it is worth bearing in mind that it is a simplification. There are more and more T_{helper} cell populations discovered. One explanation to this is that it has been shown that depending on the time, place and conditions there are T_{helper} cell types that co-express markers and even functionalities of multiple subsets and also one subset can be cross-differentiated into another subset (Caza and Landas, 2015; Geginat et al., 2014). The more the technologies for measuring and analyzing multiple markers at a time on single cell level progress, the more it has been becoming clear that the classical distinction into few subsets is only a simplified guideline (Wong et al., 2016). Instead, there is accumulating evidence against the traditional discrete classification of T_{helper} cell subsets, and the available information rather seems to support a conceptual model where the types of T_{helper} cells are continuous (as opposed to discrete) and they are not terminally fixed; this is also referred to as plasticity.

1.3 T_{helper} cells in Multiple Sclerosis (MS)

1.3.1 MS: an autoimmune disease

One of the diseases where autoreactive T cells play an important role is Multiple Sclerosis (MS). MS is an autoimmune neuroinflammatory disease, in which immune cells infiltrate the central nervous system (CNS) and cause inflammation and tissue destruction [reviewed in (Becher et al., 2016; Dendrou et al., 2015; Goverman, 2009)]. Specifically, this destruction involves demyelination – that is, destruction of the myelin sheath which normally forms the electrically insulating surrounding layer around the axons of some nerve cells, such as in the white matter of the brain. Myelin consists of different cell types, lipids and proteins including myelin basic protein (MBP), myelin oligodendrocyte glycoprotein (MOG) and proteolipid protein (PLP) [reviewed in (Mallucci et al., 2015)]. The T_{helper} cells contributing to MS are believed to recognize peptides derived from myelin protein components as a self-antigen (i.e. they are autoreactive), and corresponding self-antigens have been identified from MOG, PLP, MBP and other proteins. Also, certain viruses which share binding-motifs for MHC class II and TCR binding with MBP have been proposed as potential triggers of MS [reviewed in (Mallucci et al., 2015)]. The autoreactive T cells and other immune cells can pass the blood brain barrier in MS, that is facilitated by expression of the appropriate chemokine receptors, and in the CNS they secrete cytokines that signal to other immune cells such as macrophages, which then destroy the myelin sheath. Thus, one key feature of T_{helper} cells that are harmful in MS is the set of cytokines they produce, i.e. their cytokine profile.

1.3.2 Characterization and differentiation of pathogenic auto-reactive T_{helper} cells

The molecular and cellular mechanism of MS and the role that specific cytokines play in it can be studied by interventional experiments in mice (and rats). Specifically, knockout or conditional knockout strains and/or cell transfer experiments have been used in combination with disease induction by peptides from protein components of myelin which leads to experimental autoimmune encephalomyelitis (EAE), the rodent model of MS. Furthermore, specific T cells from TCR-transgenic mice expressing TCRs that recognize myelin autoantigens such as MBP peptides have been useful tools (Kuchroo et al., 1994). In contrast, studies of human MS are limited to observational approaches, for example by studying the relationship between genetic variation [single-nucleotide polymorphism (SNP), copy number variation (CNV)] and disease risk or disease characteristics, which usually allows for identification of associations.

A crucial element in understanding the role of T cells in MS would be to identify the cytokine(s) that confer pathogenic functions to T cells, both those cytokines acting on the differentiating T cells as well as the cytokines produced by the T cells themselves.

From studies applying EAE models in mice and from descriptive human MS case-control studies, the T cell-released cytokines that have been identified to possibly have a role in the etiology of MS and EAE are IL-17A/F, IFN-γ, IL-22 and GM-CSF (Codarri et al., 2011; Hartmann et al., 2014; Noster et al., 2014; Ponomarev et al., 2006; Ramesh et al., 2014; Rolla et al., 2014).

A multitude of studies has focused on Th17 cells as one important cell subset involved in EAE induction (Cua et al., 2003; Langrish et al., 2005; McGeachy et al., 2009), and hence *in vitro* differentiation conditions allowing polarization towards the Th17 subset have been applied. In detail, it has been shown that T_{helper} cells differentiated under Th17-polarizing conditions in the presence of IL-23 were capable of inducing EAE (i.e., are “pathogenic”), unlike those differentiated in the presence of IL-12 (Cua et al., 2003; Parham et al., 2002). Furthermore, IL-23 has been shown to be necessary for induction of IL-17A/F producing CD4+ T cells *in vivo* since IL-23 (p19) knockout mice lack IL-17A/F + cells (Langrish et al., 2005). These observations pointed to the importance of IL-23 as a central cytokine in EAE, and to murine IL-17 producing T_{helper} cells (Th17) as key T_{helper} subset in EAE (Cua et al., 2003; Langrish et al., 2005; McGeachy et al., 2009). However, mice deficient in IL-17A or IL-17F were still susceptible to EAE (Haak et al., 2009), suggesting that other properties than IL-17 expression, but correlating with the EAE-inducing pathogenic “Th17” phenotype, may be involved in pathogenicity.

IL-23 is a cytokine not produced by T cells, and the above findings were especially interesting because the heterodimeric proteins IL-12 and IL-23 share the p40 subunit, and hence the later discovered IL-23 may be the explanatory factor for divergent outcomes in clinical studies targeting the different subunits of IL-12 in human immune diseases (Teng et al., 2015).

In mice, IL-23, TGF-β3 and IL-1β have been shown to act as induction factors of pathogenic "Th17" cells, although such studies require in-depth analysis for accurate interpretation since the given cytokines have never demonstrated the effect alone, only in combination with other cytokines suggesting that their effects are conditional (El-Behi et al., 2011; Ghoreschi et al., 2010; Kara et al., 2015; Langrish et al., 2005; Lee et al., 2012; Wu et al., 2013). It has also been suggested that a modest increase in sodium chloride (NaCl) concentration in the cell culture medium towards physiological levels is able to induce pathogenic Th17 cells by activating the salt-sensing kinase SGK1, which in turn increases IL-23R expression, thereby enhancing the IL-23R signaling pathway (Kleinewietfeld et al., 2013; Wu et al., 2013). TGF-β3 has been proposed to be endogenously secreted as a result of IL-23R signaling, and subsequently conferring pathogenic properties on the developing Th17 cells (Lee et al., 2012). However, this is yet to be confirmed, since another independent study has shown that Th17 cells induced in the presence of either TGF-β1 or TGF-β3 were not pathogenic (Lee et al., 2015).

Given the above experimental results and considering that both non-pathogenic and pathogenic (EAE-inducing) Th17 cells express IL-17 in this animal model, the question arises: what makes a "Th17" cell pathogenic? In other words, even though each of the Th17 cell populations induced under different Th17-polarizing conditions produced IL-17, not all of them were able to induce EAE in cell transfer experiments. Furthermore, complete or T cell-specific deficiency of IL-17 did not prevent EAE induction (Codarri et al., 2011; Haak et al., 2009). Together, these observations suggest that IL-17 itself is not sufficient for disease induction.

To understand the difference between the pathogenic and non-pathogenic "Th17" populations, the transcriptome of Th17-polarized cells induced by different corresponding cytokine cocktails has been studied, showing that pathogenic and non-pathogenic populations expressed different sets of genes. Consequently, signature genes associated with pathogenic or non-pathogenic Th17 cells have been suggested (Gaublomme et al., 2015; Ghoreschi et al., 2010; Lee et al., 2012; Wu et al., 2013).

One of the genes associated with a pathogenic signature is the gene *Csf2* (Colony Stimulating Factor 2) encoding for the cytokine granulocyte-macrophage colony stimulating factor (GM-CSF). When added during *in vitro* differentiation of murine Th17 cells, IL-1β and IL-23 induced GM-CSF^{+}IL-17^{+} double-positive cells (El-Behi et al., 2011). Similarly, Th17 cells differentiated in the presence of TGF-β3 or elevated NaCl concentrations displayed increased *Csf2* expression on the population level, but have not been studied at single cell resolution (Lee et al., 2012; Wu et al., 2013). A more recent study has used single cell RNA-sequencing to analyze the heterogeneity of murine Th17 cells isolated *ex vivo* from the lymph nodes or CNS of mice at the peak of EAE disease, or differentiated *in vitro* under pathogenic (IL-23 + IL-1β + IL-6) or non-pathogenic (TGF-β1 + IL-6) Th17 conditions (Gaublomme et al., 2015). Here, sorted IL-17+ cells displaying a Th1-like memory cell phenotype from the CNS up-regulated *Csf2* mRNA (Gaublomme et al., 2015). Interestingly and against the notion of IL-23, SGK1 and GM-CSF being crucial in pathogenic Th17 differentiation, IL23R- or SGK1-deficient T cells differentiated under Th17-polarizing conditions expressed increased *Csf2* levels compared to the wild-type counterpart (Wu et al., 2013).

Altogether, these studies shed light on the different cell populations involved in EAE, but due to the overlapping patterns of the different signature cytokines from the diverse Th cell subsets it is difficult to conclude which cytokines are actually necessary for pathogenicity and which ones are just associated with the pathogenic phenotype, either on single cell level or even only on the population level due to similar inducing factors. To study which cytokines are necessary in EAE induction, key experiments were those with cytokine-knockout models, which showed that deficiency of either IL-17A, IFN-γ, or GM-CSF specifically in T cells impaired their pathogenic potential, at least to some extent (Codarri et al., 2011).

1.3.3 GM-CSF as a crucial cytokine in MS

As described above, there are several cytokines that have been implicated in MS based on studies using knockout and other approaches, but the one that really seems to stand out is GM-CSF encoded by the gene *CSF2*. GM-CSF has been shown to be necessary for EAE induction in mice, since mice lacking the gene encoding GM-CSF (*Csf2*) do not develop EAE upon induction, unlike their wild-type counterpart (McQualter et al., 2001). In fact, GM-CSF deficiency led to complete resistance to EAE induction. In contrast, lack of IL-17A, IL-17F or IFN-γ had no or only marginal effects on susceptibility to EAE disease induction (Ferber et al., 1996; Haak et al., 2009).

It is worth noting though that these experiments might be confounded by the fact that GM-CSF is also an important factor for the development of the myeloid lineage, therefore these mice might have an abnormal immune system, and it might be more advantageous to use conditional knockout strains such as those lacking GM-CSF specifically in T cells. Similar issues of potential systemic effects apply to knockout mice for other cytokines. An alternative to the use of conditional knockout strains is to perform adoptive transfer experiments with T cells that are deficient in the cytokine of interest into cytokine-sufficient mice, and in this context, TCR transgenic T cells recognizing myelin-derived auto-antigens are particularly useful for studying EAE. Confirming the role of T cell-derived GM-CSF, it has been demonstrated in mice by cell transfer experiments with cytokine knockout T cells that GM-CSF produced in particular by autoreactive T cells is necessary for the onset of EAE (Codarri et al., 2011; Ponomarev et al., 2006). Furthermore, it was demonstrated that GM-CSF production by "Th1" and "Th17" cells is required for their encephalitogenicity (El-Behi et al., 2011). Importantly, INF-γ as well as IL-17A production by T cells was dispensable, while GM-CSF production was necessary for induction of EAE (Codarri et al., 2011; Ponomarev et al., 2006). Along the same lines, murine T cells polarized to produce GM-CSF induced more severe EAE upon adoptive transfer compared those polarized to produce IFN-γ or IL-17, although the latter also induced disease (Codarri et al., 2011). However, overlapping cytokine profiles of such *in vitro* polarized cells cannot be excluded.

Despite the important insights that have been derived from above-described animal studies with the EAE and other models, it remains debated to which extent these results reflect the situation in human MS, since obviously experiments to test the pathogenicity of certain cell populations are not possible in human.

Regarding the human biology, from observational experiments GM-CSF seems to be associated with MS since it has been shown that at the site of disease, in the cerebrospinal fluid, there are increased frequencies of GM-CSF producing T_{helper} cells in MS patients compared to controls (Noster et al., 2014). The fraction of GM-CSF producing CD4+ T cells was not elevated in peripheral blood of MS patients in the same study (Noster et al., 2014), although this was observed in another report (Hartmann et al., 2014). IFN-γ production in CD4+ T cells showed the same pattern as GM-CSF, with increased fractions of IFN-γ+ cells in cerebrospinal fluid of MS patients compared to controls, while the fraction of IL-17+ cells did not significantly differ between the two groups in the same study (Noster et al., 2014). Notably, myelin-reactive T cells have been detected in both MS patients as well as healthy donors with comparable frequencies, but have been determined to produce more IFN-γ, IL-17 and GM-CSF and at the same time less IL-10 in MS patients compared to healthy donors, in both oligoclonal populations or single cell-derived clones of

myelin-reactive T cells (Cao et al., 2015). While Th17 cells producing GM-CSF have been proposed as pathogenic cells in EAE, in the human MS studies, GM-CSF rather seems to associate with Th1-like than with Th17-like cells. Although differences in murine and human CD4+ subset differentiation and classification emerge, rodent models have been used to enhance the understanding of which T cells are involved in the pathogenesis of EAE, which may or may not reflect the scenario in MS depending on the specific conditions and cells under study.

Interestingly, enhanced GM-CSF levels including elevated frequencies of GM-CSF producing CD4+ T cells have also been observed in another human autoimmune disease at the site of inflammation – that is, synovial fluid of patients with juvenile arthritis (Piper et al., 2014; Wicks and Roberts, 2015). According to an assumed important role of GM-CSF in these diseases, several ongoing clinical studies are based on neutralizing GM-CSF in MS or arthritis (Wicks and Roberts, 2015).

1.3.4 Induction and phenotype of GM-CSF producing CD4+ T cells

As described above, GM-CSF produced by autoreactive T cells seems to play an important role in the pathogenesis of MS. But how does GM-CSF fit into our view of classical T_{helper} cell subsets? There are several T_{helper} cell subsets, each one with a specific cytokine profile and function, and each one differentiates when a certain cytokine environment is present (see paragraph 1.2, Figure 2), but it remains unclear how to position GM-CSF producing CD4+ T cells in the T_{helper} subset space.

GM-CSF production has been detected in cells that may be classified to other Th subsets (such as those co-expressing IFN-γ, IL-3, or IL-17). As described above, IL-17 and GM-CSF can be co-expressed in murine T cells (El-Behi et al., 2011) but interestingly, their expression was mutually exclusive in human CD4+ T cells at least in one study (Noster et al., 2014). Here, GM-CSF can be co-expressed with other lineage-defining cytokines such as IFN-γ, but interestingly GM-CSF single-positive ("GM-CSF only") human CD4+ T cells that do not express other classifying cytokines or lineage-defining transcription factors have also been observed (Noster et al., 2014; Zielinski, 2014). However, whether the GM-CSF producing T cells are a specific subset of T cells, and if so which markers define it, is unclear to date.

In vitro differentiation of T cells has been used as an experimental system aiming to gain knowledge about factors regulating GM-CSF production in human CD4+ T cells, with partially conflicting results and several open questions (summarized in Figure 3). For example, TGF-β1 or TGF-β3 were found to decrease GM-CSF production in human CD4+

T cells in one study (Hartmann et al., 2014), while TGF-β1 had no effect in another (Noster et al., 2014). IL-23 and IL-6 did not augment GM-CSF (Hartmann et al., 2014; Noster et al., 2014), whereas IL-2 or IL-7 signaling induced GM-CSF expression in a STAT5-dependent manner, and IL-1β induced IFN-γ^+ GM-CSF$^+$ double-positive cells (Hartmann et al., 2014; Noster et al., 2014).

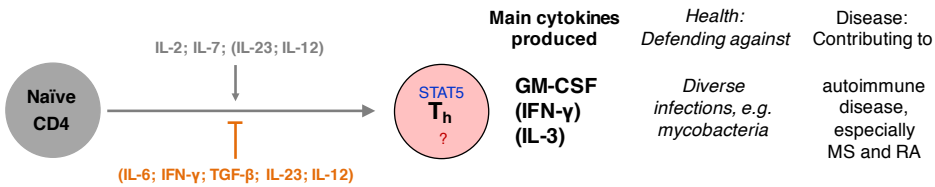

Figure 3: Factors known to influence the differentiation of GM-CSF producing T_{helper} cells, and the transcription factors, cytokines and functions linked to them. The cytokines given in grey have been described to enhance differentiation, while factors in orange have been described to inhibit the differentiation of GM-CSF producing T_{helper} cells. Cytokines in parentheses showed the indicated effect in only some but not all studies, and/or differ between mouse and human. In blue, the STAT family member (STAT5) reported be involved is shown. Question mark indicates that no "master transcription factor" has been identified for GM-CSF producing T_{helper} cells. The cytokines described to be co-produced with GM-CSF are shown in black. T_{helper} cell produced GM-CSF has been reported to play a role in protection against diverse infections and in the pathology of autoimmune disease, such as multiple sclerosis (MS) and rheumatoid arthritis (RA).

Altogether, the above studies imply that GM-CSF produced by CD4+ T cells is a crucial contributor to MS and EAE. So far it has not been clearly defined whether GM-CSF producing T cells are a separate subset and if so, what their lineage-defining transcription factor is or which other markers may enable the identification of these cells. Understanding the generation and characteristics of GM-CSF+ CD4+ T cells may be important to improve treatment options for patients with autoimmune diseases in the future.

2 AIMS

The overall aim of this thesis was to characterize the differentiation and molecular properties of human T_{helper} cell subsets that contribute to multiple sclerosis (and other autoimmune diseases) by combining experimental and computational approaches. Here, the focus was especially on those T_{helper} cells producing the cytokines GM-CSF, IFN-γ, and/or IL-17.

The specific aims per paper were the following:

Paper I

— Identify promoting and inhibitory cytokines and other environmental signals for the *in vitro* differentiation of human GM-CSF producing T_{helper} cells from naïve CD4+ T cells

— Quantify the similarity of human GM-CSF producing T_{helper} cells to other T_{helper} cell subsets, both on single cell and population level using computational flow cytometry

— Define subpopulations of human GM-CSF producing T_{helper} cells and map them to the inducing cytokine signals

Paper II

— Explore the regulation & function of FOXP3 isoforms in the differentiation of pro-inflammatory T_{helper} subsets

Paper III

— Develop a user-friendly web-based platform that allows the application of a wide range of machine learning and causal discovery methods on flow and mass cytometry data (some of these methods have been applied in Paper I programmatically)

Paper IV

— Identify the defining characteristics of human GM-CSF producing T_{helper} cells as well as memory T_{helper} cells *ex vivo* in terms of both gene expression and chromatin accessibility

— Reconstruct a directed gene regulatory network representing the transcription factor binding events that define the molecular profile of human GM-CSF producing T_{helper} and memory T_{helper} cells

— Assign pathways and diseases that are associated with the profile of human GM-CSF producing T_{helper} cells and memory T_{helper} cells

3 RESULTS

3.1 Computational flow cytometry studies on the differentiation of pro-inflammatory T_{helper} subsets

Both GM-CSF and IL-17 producing T_{helper} cells are pro-inflammatory T cell subsets, and they both have been implicated in autoimmune diseases, such as MS and RA. In Paper I and Paper II, I studied the differentiation of these T_{helper} cell subsets from different aspects (discussed below). Paper I involved analysis of large amounts of flow cytometry data and it is also concerned with application of advanced statistical methods on cytometry data, while Paper III is concerned with the development of a user-friendly platform for such tasks.

3.1.1 Factors affecting GM-CSF producing T_{helper} cell differentiation – Paper I

3.1.1.1 Cytokine effects

GM-CSF producing T_{helper} cells have been associated with MS and EAE, but it has not been clearly determined which cytokines induce or repress their differentiation. There are few studies on this question in the literature, and even among that few, some of the conclusions are contradictory, which may be explained by differences in terms of species or experimental system. For example, IL-23 has been shown to promote the differentiation of GM-CSF producing T_{helper} cells in mice (El-Behi et al., 2011) but repress it in human *in vitro* experiments (Noster et al., 2014).

Therefore, in this part of the thesis the following question was addressed: What cytokines promote or repress the differentiation of human GM-CSF producing T_{helper} cells? The approach here was to use *in vitro* T cell differentiation assays and data-driven analysis of cytokine effects. For this purpose, freshly isolated naïve CD4+ T cells were cultured in the presence of anti-CD3/CD28 coated activating beads and various cytokines (that were previously described to influence the differentiation of T_{helper} subsets) or combinations thereof as "cytokine cocktails". After 5 days (unless otherwise indicated), GM-CSF as well as other markers (for T_{helper} subsets and activation) were quantified by flow cytometry. Importantly, in this study all the cytokine effects were measured under the same experimental conditions, therefore they were directly comparable.

In order to quantify the GM-CSF promoting or inhibitory effects in human T_{helper} cells that can be attributed to specific cytokines and cytokine pairs, least absolute shrinkage and selection operator (LASSO) linear regression was applied. This type of regression is advantageous to prevent overfitting since it includes regularization and it is also useful for selection of the simplest and most accurate model by performing cross-validation. After cross-validation one can inspect the mean cross-validated error as a function of the regularization parameter (λ or $\ln(\lambda)$) and choose the regularization parameter value (thereby the model) with the smallest mean cross-validated error.

This analysis showed that TGF-β strongly induced the differentiation of GM-CSF producing T_{helper} cells, especially when combined with IL-1β. On the other hand, IL-23 had a repressing effect. This was conceptually unexpected since GM-CSF has been described to contribute to inflammatory diseases and classically TGF-β is referred to as an anti-inflammatory cytokine, although context dependent effects have been described [reviewed in (Massagué, 2012; Morikawa et al., 2016)], and IL-23 is referred to as a pro-inflammatory cytokine. The inhibitory effect of IL-23 on GM-CSF expression observed herein was in line with a study on human T_{helper} cells (Noster et al., 2014) but not with a different study in mice (El-Behi et al., 2011).

3.1.1.2 Context dependency

In human, TGF-β was shown to inhibit GM-CSF expression in T_{helper} cells in one study (Noster et al., 2014), it had no significant effect in another study (Hartmann et al., 2014), and in the present study it induced the expression of GM-CSF in human T_{helper} cells. Therefore, to find an explanation for this discordance, differences in terms of experimental conditions (protocol) were investigated and compared. An important difference between this and the other studies was the type of activation. In addition, sodium concentration has been previously described to influence the differentiation of T_{helper} cells (Kleinewietfeld et al., 2013; Wu et al., 2013). Therefore, the effect of these factors was further investigated.

In our experiments, bead-bound anti-CD3/CD28 activation was used, whereas in the studies of (Hartmann et al., 2014; Noster et al., 2014) plate-bound anti-CD3 and soluble anti-CD28 based activation was used. Therefore, these two types of *in vitro* activation were compared in the present work with regards to their effect on GM-CSF expression in activated human T_{helper} cells, in the presence or absence of TGF-β. In these experiments, with anti-CD3 and soluble anti-CD28 based activation TGF-β showed an inhibitory effect on GM-CSF expression, which matched the observation described by (Noster et al., 2014). Therefore, activation type might be one factor that can explain the different effects of TGF-β observed on GM-CSF expression.

The other factor further tested was the sodium chloride (NaCl) concentration in the medium, which has been described to modulate T_{helper} cell differentiation (see section 1.3.2, (Kleinewietfeld et al., 2013; Wu et al., 2013)). Interestingly, my results showed that TGF-β had a promoting effect on the differentiation of GM-CSF producing T_{helper} cells in the presence of physiological sodium chloride concentration, and a repressing effect in the presence of low sodium chloride concentration.

Since TGF-β affected GM-CSF expression in human T_{helper} cells differently, depending on what type of activation and what concentration of sodium chloride was present, a possibility to consider was that the activation type and/or sodium chloride concentration modulates the total percentage of GM-CSF producing T_{helper} cells only by influencing the extent of proliferation, thus indirectly. To control for this and to determine whether this was the case, throughout the experiments investigating the effect of activation type and sodium chloride concentration, cells were labelled with a proliferation dye and the percentage of GM-CSF producing T_{helper} cells was analyzed conditioning on the cumulative number of cell cycles. This showed that the differential effect of TGF-β, depending on activation type or sodium chloride concentration, was independent of proliferation. In addition, the effect of TGF-β on GM-CSF expression was investigated in the presence or absence of a STAT5 inhibitor. Conditioning on the cumulative number of cell cycles, the results showed that the GM-CSF promoting effect of TGF-β (with bead-bound anti-CD3/CD28 activation, and physiological sodium chloride concentration) was independent of the STAT5 signaling pathway.

These results exemplify the possible context dependent effects of cytokines, and emphasize the importance of experimental design.

3.1.2 Positioning GM-CSF producing T_{helper} cells in the T_{helper} space – Paper I

It has not been clearly determined according to the available literature whether GM-CSF producing T_{helper} cells are a distinct T_{helper} cell population and to which other subsets they are most similar to (see section 1.3.4). Therefore, these questions are addressed in the following sections.

3.1.2.1 GM-CSF as an individual cytokine

First, the pairwise Spearman correlation was investigated between GM-CSF as a cytokine and several other main T_{helper} subset markers in two different ways: (i) on the population level across stimulation conditions (here stimulation conditions be thought of as "perturbations"), and on the single cell level, i.e. across single cells, treating the stimulation conditions as replicates. Interestingly, the results differed in the two different cases: FOXP3 was highly correlated with GM-CSF on the population level across conditions, and IFN-γ was highly correlated with GM-CSF on single cell level under the different conditions. This suggested the possibility that GM-CSF and FOXP3 might be induced by similar cytokine signals in human differentiating T_{helper} cells but on the intracellular pathway level they might be rather independently regulated. On the other hand, IFN-γ might be co-regulated with GM-CSF on the intracellular pathway level, since under various stimulation conditions their expression co-varies across single cells.

3.1.2.2 GM-CSF producing T_{helper} cells as a cell population

As part of assessing the similarity of GM-CSF producing T_{helper} cells to other T_{helper} subsets, next, the composition of the GM-CSF producing T_{helper} cell population was investigated. Here, the aim was to determine: (i) whether they are a homogenous population, (ii) whether different stimuli give rise to different subsets within GM-CSF producing T_{helper} cells, and (iii) to what extent these subsets express the other main T_{helper} subset markers included in the analysis.

It is worth noting that the semantics of what we mean by a subset (or "cell population") is not obvious, and it is useful to explicitly clarify this. Although defining a cell population as "cells with identical function" might be most appropriate in a biological sense, here we define cell populations purely in a data-driven way as "clusters" of cells based on their multivariate distribution.

To address the above questions, t-SNE dimensionality reduction and model based clustering was used to identify and analyze subgroups of GM-CSF producing T_{helper} cells.

This analysis showed, that GM-CSF producing T_{helper} cells are heterogeneous with respect to other T_{helper} subset markers and that the population composition depends on the cytokine environment since it differed in the presence or absence of TGF-β. For example, in the presence of TGF-β there was a subpopulation present resembling to Tregs, showing high FOXP3 and CD25 expression, and low expression of IL-2. However, these are necessary but not sufficient hallmarks of Tregs, and in human, FOXP3 can also be induced simply by T cell activation (Pillai et al., 2007; Wang et al., 2007).

These results together showed that GM-CSF producing T_{helper} cells can be induced in multiple different ways and that the composition of the cell population is dependent on the mode of induction. Therefore, conceptually relating the process of T_{helper} cell differentiation from single cell to population level seems complex, and in the future, definition and experimental testing of theoretical frameworks might help to better understand this relationship (Hong et al., 2012).

3.1.3 Regulation & function of FOXP3 isoforms in T_{helper} cells – Paper II

FOXP3, known as the master regulator transcription factor of Tregs, exists in several alternative splice isoforms (in human but not in mouse) and the role of this diversity is not completely understood (Aarts-Riemens et al., 2008). It has been shown that the FOXP3 isoform lacking exon 2 and exon 7 is not fully functional (Mailer et al., 2009). Initial experiments from this study suggested that Crohn's disease patients expressed higher levels of the FOXP3 isoform that lacks exon 7, compared to healthy donors, without the total FOXP3 expression being significantly different between the groups. Therefore, we wanted to study further the reasons and consequences of FOXP3 isoform differences.

In order to identify cytokines that modulate alternative splicing of FOXP3, we chose to activate Tregs in the presence of IL-1β, IL-6 or TNF-α, since FOXP3 naturally occurs in Tregs and these cytokines are key pro-inflammatory cytokines, thus possibly having a role in shifting the balance of immune regulation. Of these cytokines, IL-1β was able to increase the expression the FOXP3 isoform lacking exon 7 in activated Tregs.

An interesting connection might be that in Paper I IL-1β alone but especially IL-1β together with TGF-β (that is known for its FOXP3 inducing effect), also promoted the differentiation of GM-CSF producing T_{helper} cells. It might be interesting to see in the future whether these induced GM-CSF producing T_{helper} cells have any alteration in the balance of FOXP3 isoforms.

IL-17 has been linked to several autoimmune diseases, as well as GM-CSF (that is studied in Paper I and Paper IV). IL-2 is a cytokine generally important for T cell activation. In order to identify cytokines that are modulated by FOXP3 alternative splicing, we induced alternative splicing of FOXP3 by transfecting Tregs with morpholino antisense oligonucleotides (MAO) that prevent binding of splice-directing small nuclear ribonucleoproteins to the exon-exon boundaries of FOXP3 pre-mRNA [reviewed in (Kole et al., 2012)], and measured the percentage of IL-2 and IL-17 producing cells by flow cytometry. These experiments showed that removal of exon 7 in Tregs increased the percentage of IL-2 producing cells, and removal of both exon 2 and exon 7 increased the percentage of both IL-2 and IL-17 producing cells.

In order to test whether there is an association between altered FOXP3 isoform levels and IL-17 levels in a disease setting, we quantified the relative expression of FOXP3 isoforms and IL-17, and analyzed the correlation between them using colon biopsies from Crohn's disease patients. Quantified by rank based Spearman correlation, IL-17 was significantly

correlated with the FOXP3 isoform lacking exon 7 in affected tissue from Crohn's disease patients.

This study can serve as an example showing how the regulation and balance of pro- and anti-inflammatory processes are interconnected.

3.1.4 A web based tool for computational analysis of single cell cytometry data – Paper III

Throughout all the studies in this thesis, and especially heavily in Paper I, flow cytometry was used to measure responses and behavior of T cells on the single cell level. Flow cytometry and its newer relative mass cytometry can be extremely valuable tools due to their ability to measure (i) multiple variables, (ii) on the single cell level, (iii) in a large number of cells. But in practice, they are only as valuable as the extent to which we are able to extract and interpret information from the data they produce. Exactly due to their high dimensional nature and their large sample sizes (number of cells), often combined with complex experimental design (for example time course with multiple groups and interventions, etc.), flow- and mass-cytometry data can be hard to interpret by eye. To draw conclusions from such data, various machine learning approaches can be useful (regression, dimensionality reduction, clustering to name a few). An example of how these methods can be used for flow- and mass-cytometry data is Paper I, but that is by no means a comprehensive example and there are far more possibilities of analysis depending on the question and the design of the experiments.

Since application of machine learning or generally advanced statistical methods usually requires programming skills, let alone the theoretical background, these methods are usually not accessible to wet-lab biologists. Therefore, in Paper III we created a tool, that gives the possibility of applying such methods, to wet-lab biologists via a user-friendly interface, without the need for programming skills. Thus, in collaboration with computer science and data science experts, we created SCENERY (Single CEll NEtwork Reconstruction sYstem).

SCENERY is web application that allows the users to upload cytometry data in FCS or text file (comma or tab separated) format, and to analyze this data. The different types of analysis are divided to three categories: pre-processing, statistical analysis, and network reconstruction. The structure of SCENERY is modular, therefore users can extend the set of the available analysis methods with new ones, given that the input/output of the new method complies with the standards of the web server.

3.2 Molecular profile & gene regulatory network of *ex vivo* memory & GM-CSF producing T_{helper} cells – Paper IV

The aim in this part of the thesis work was to get a highly detailed picture and understanding of what makes a GM-CSF producing T_{helper} cell by studying their chromatin and transcriptome together, since this might be important due to their relevance in MS (see section 1.3.2 – 1.3.3). It was apparent from previous parts of the thesis that (i) GM-CSF producing T_{helper} cells can be induced in different ways, (ii) they are heterogeneous, and (iii) their expression properties depend on the induction conditions. Therefore, to characterize the gene regulation and expression profile of GM-CSF producing T_{helper} cells, I chose to use a more physiological system using *ex vivo* isolated GM-CSF producing T_{helper} cells. Physiological relevance was particularly important in this case since this study involved measurement and in-depth analysis of thousands of variables (genes and chromatin regions).

The main contrast of interest was to compare GM-CSF producing *vs.* non-producing T_{helper} cells to each other. GM-CSF producing T_{helper} cells were isolated from total (bulk) CD4+ T cells by using a cytokine capture assay that contains an antibody construct that is immobilized to the cell surface and binds the secreted cytokine, and because the construct also contains metal beads, the cells can subsequently be isolated by magnetic bead isolation. This procedure yielded three cell populations: bulk, GM-CSF producing, and non-producing CD4+ T cells. The bulk population can serve as a control to estimate the effects on chromatin and transcriptome purely arising due to the GM-CSF capture assay procedure (that involves the use of a magnetic column), since the bulk population mostly consisted of GM-CSF non-producer cells and therefore it is ought to be very similar to the GM-CSF non-producer (negative) fraction, except that the latter one has been through the capture assay procedure.

In addition, memory and naïve T_{helper} cells were used as controls for the following reasons: Memory T_{helper} cells are more likely to produce GM-CSF (and cytokines in general) than naïve T_{helper} cells. This means that when comparing the chromatin and transcriptome of GM-CSF producing *vs.* non-producing T_{helper} cells to each other, some of the differences inherently originate from the memory *vs.* naïve status. Therefore, to be able to account for the differences due to memory *vs.* naïve status, and separate these from the differences that arise due to GM-CSF producing *vs.* non-producing status, two additional cell populations were isolated and analyzed: memory and naïve T_{helper} cells. Also, apart from serving as controls, the comparison of memory *vs.* naïve T_{helper} cells is also interesting by itself, due to the fact that we they have not been studied before using chromatin and transcriptome data together.

The read-outs used in this study were ATAC-seq and RNA-seq to study the chromatin activity and gene expression respectively. The main comparisons to make were GM-CSF producing *vs.* non-producing (GM-CSF contrast) and memory *vs.* naïve T_{helper} cells (memory contrast). The aim was to integrate the two data types (ATAC-seq and RNA-seq) and to propose an explanation of the gene expression changes observed by RNA-seq, by identifying putative transcription factor binding events in active chromatin regions using the ATAC-seq data, and thus creating two gene regulatory networks corresponding to the memory and GM-CSF contrasts.

3.2.1 Transcriptome analysis: differential expression & disease enrichment

In order to identify differentially expressed genes (as well as differentially accessible chromatin regions), generalized linear models were used (based on the negative binomial distribution, edgeR (McCarthy et al., 2012; Robinson et al., 2010)). In these models, the independent variables were coded as dichotomous variables using the levels of cell types (GM-CSF-, GM-CSF+, bulk, naïve and memory CD4+ T cells) and the levels of the biological replicates, independently. The contrasts of interests extracted were the GM-CSF and the memory contrasts (defined above), and as a result of the model definition, the effects were adjusted for biological variability.

Overall, based on the differential expression results and also based on dimensionality reduction (PCA & t-SNE), the memory contrast showed larger differences than the GM-CSF contrast in terms of gene expression (of note, this was less pronounced on the level of chromatin accessibility measured by ATAC-seq). Among the differentially expressed genes, there were several with a described function in T cells. At the same time, an interesting set (~9%) of those genes differentially expressed in the GM-CSF contrast was unique this contrast (and was not differentially expressed in the memory contrast), and this contained several genes with unexplored function in T cells.

When performing a rank based gene set enrichment using the combined measure of the significance and direction of differential expression in the GM-CSF contrast (-log10(FDR) × sign(log2(Fold Change))) and gene sets corresponding to diseases, the significantly enriched diseases contained mostly immune diseases. Some of these, such as MS and RA, have been previously linked to GM-CSF producing T_{helper} cells. Mycobacterial infections were also significantly enriched diseases and the role of GM-CSF and GM-CSF producing T_{helper} cells has been described in these diseases (Chroneos et al., 2009; Denis and Ghadirian, 1990; Rothchild et al., 2017). Furthermore, several of the genes differentially expressed in this study in the GM-CSF contrast also showed altered expression between myelin reactive *vs.* non-reactive T_{helper} cell clones in MS patients but not in healthy controls

in the data from another study (Cao et al., 2015). The enrichment of several expected diseases was indicating that the data points to the right direction and that it is worth further exploration in a data driven way, possibly leading to unexpected conclusions as well.

3.2.2 Prediction of transcription factor binding

The chromatin activity was studied by a method called ATAC-seq. In brief, it is a method that is based on the idea that active chromatin regions are accessible, and these regions can therefore be cut by a transposase enzyme called Tn5 (Buenrostro et al., 2013). Thus, this method allows identification of accessible chromatin regions and therefore the study of chromatin activity and its relation to gene regulation.

In this study, the main purpose of using ATAC-seq was to identify putative transcription factor binding events in active chromatin regions for gene regulatory network reconstruction. In order to identify putative transcription factor binding events three main steps are necessary: peak-calling to identify accessible chromatin regions, footprinting in accessible regions to identify sites where binding of a transcription factor is likely, and motif scanning in the footprints to make a statistically educated guess regarding which transcription factor is a likely binder in the footprint. By applying this procedure, a genomic region can be assigned as a predicted target of a transcription factor. The genomic region can be assigned to a gene as well (by several possible algorithms). Therefore, this way a connection (directed edge) can be drawn from a transcription factor to a gene, and this is the basis for the directed network reconstruction.

3.2.3 Data integration: directed network reconstruction from chromatin & transcriptome information

In order to be able to conclude more from the ATAC-seq and RNA-seq data together, than from the information they independently hold, the aim was to combine, i.e. integrate these two data types. The main approach in this study to integrate the ATAC-seq (chromatin) and RNA-seq (transcriptome) data was a directed network reconstruction procedure based on the following notions: (i) ATAC-seq data can be used to identify putative transcription factor binding events (i.e. directed edges, as briefly described above); (ii) RNA-seq data can be used to quantify both the relative abundance and the differential expression of genes, which allows filtering of non-detectable, non-differential or lowly expressed genes, if justified.

The directed network reconstruction was carried out both for the GM-CSF contrast ("GM-CSF network") and also for the memory contrast ("Memory network") and the networks were analyzed in side-by-side fashion. The reconstruction procedure consisted of two main steps: identification of source nodes, and identification of target nodes.

3.2.3.1 Identifying source nodes

This step was based on the idea that important transcription factors (source nodes) are likely to bind a high fraction of differentially accessible peaks, in other words they have a global effect. Therefore, peaks were ranked based on their combined measure of significance and direction of differential accessibility (-log10(FDR) × sign(log2(Fold Change))), resulting in a ranked list of peaks. The peaks that contained footprints were scanned for transcription factor binding motifs, resulting in transcription factor binding sites along the ranked peak list. In order to identify transcription factors with binding sites significantly enriched in differentially accessible peaks, a normalized enrichment score was calculated (as in gene set enrichment analysis (GSEA), (Subramanian et al., 2005)), and random sampling was performed on the ranked peak list to generate a null distribution and examine how probable it would be to obtain at least the observed enrichment by chance (p value) (Sergushichev, 2016). After multiple test correction, transcription factors were selected based on their FDR values (FDR < 0.05 cutoff) and they were ranked based on their normalized enrichment score. With this approach, approximately two dozen key transcription factors were identified for both the GM-CSF and memory contrasts.

3.2.3.2 Identifying target nodes

The next step in the directed network reconstruction was to select target nodes. Target nodes in this case are peaks, with a gene assigned to them. The selection criteria were that the peak has to (i) contain a footprint with a binding motif of the source node (transcription factor) and the peak has to be (ii) either differentially accessible and/or the gene assigned to it has to be differentially expressed.

3.2.3.3 Node & network analysis

Before investigating the functional impact of transcription factors (source nodes) based on the network structure, there are a few properties of the transcription factors (based on the data) that are worth considering, since they might be necessary for the transcription factors to play an important role. These are the relative abundance and differential expression of transcription factors. The notion is that the transcription factor might have to be present at

least in a minimum quantity and it may (or may not) be important that it is differentially expressed in order to have a large impact in the network. Interestingly, the majority of the key transcription factors identified herein were not differentially expressed (although the exact fraction differed in case of the GM-CSF and Memory networks), meaning that even though they seem to be important regulators, they could not be identified by RNA-seq only on the basis of differential expression. The other characteristic to consider for key regulator candidates is their expression level, since they are less likely to be able to have an effect if they are very lowly expressed. For this purpose, the relative abundance was quantified by using the average of the logarithm of count per million reads across samples, and the distribution of expression level had a bimodal distribution, resulting in categories of "lowly" and "highly" expressed genes. Most of the identified key transcription factors were in the "highly" expressed category but not all. This information might be considered in downstream steps and, for example, for design of studies where transcription factors are followed-up.

Next, the importance of transcription factors (source nodes) can be also evaluated based on the network structure. For this type of analysis there are several measures that can be computed for each node in the network, and they elucidate the role of the node in the network from a different aspect. Therefore, in several cases it might be a good idea to choose a set of measures suited for the type of network and question, compute these multiple measures, look at the agreement and disagreement between them, and draw conclusions from them together rather than considering only one type of measure.

In this study, the networks in question were directed and for this type of networks one useful measure per node is the PageRank (Page et al., 1999). The intuitive interpretation of this measure is that a node is considered "important" (high PageRank value) if by starting on randomly chosen nodes of the network, and following the directed edges, there is a high chance of ending up on the given node. Since this rather focuses on the importance of nodes as targets, and at this step of this study the main purpose was to quantify the importance of transcription factors (source nodes), the PageRank of the network nodes was computed after inverting the directionality of all edges in the network (only for the purpose of this computation). After this computation, the nodes with high PageRank values (between the 99^{th} – 100^{th} percentile of all node values from one network) were selected from both the GM-CSF and Memory network, and investigated across the two networks. This showed, that among the important transcription factors, there are three types based on whether they are important for the GM-CSF, Memory or both networks. This approach helped to narrow down and identify the set of transcription factors that might play an important role specifically in the molecular profile of GM-CSF cells.

To study the redundancy and synergy of transcription factors in the networks, their pairwise co-binding was quantified by computing the binary distance using the binding along peaks as binary vectors. This showed that there are clear groups (clusters) of transcriptions factors that act on largely shared genomic regions. Among these groups there were both examples for groups containing subtypes of the same or different transcription factor families. Furthermore, the target genes of individual or groups of transcription factors can be studied to look for enrichment of pathways among target genes.

In summary, the above analysis, applying methodologies for data integration, network reconstruction and network analysis, allowed the identification of candidate genes for future follow-up studies with therapeutic aims.

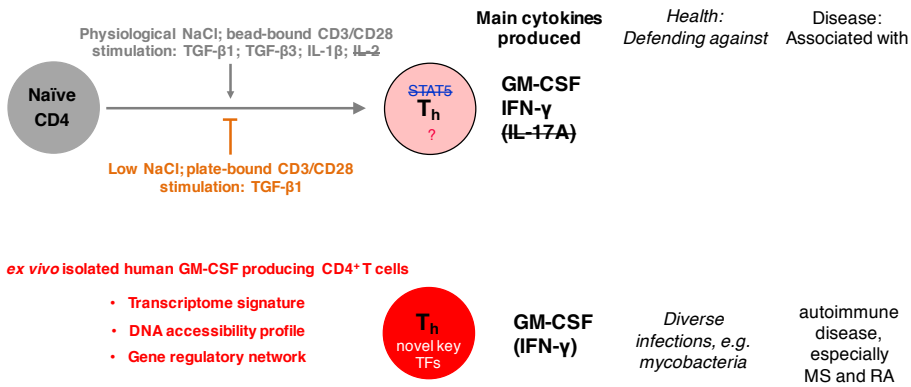

Figure 4. Summary of results on the differentiation and *ex vivo* phenotype of human GM-CSF producing T_{helper} cells. The upper part summarizes the results on the induction, *in vitro* differentiation and cytokine profile of human GM-CSF producing T_{helper} cells. The lower part summarizes the results of the in-depth characterization and analysis of the *ex vivo* phenotype of human GM-CSF producing T_{helper} cells.

4 CONCLUSIONS

The research leading to this thesis was carried out with two aspects in mind, and the aim that these two aspects have a synergistic effect on each other: (i) addressing biological questions and at the same time (ii) explore the advantages of combining different experimental and computational methodologies.

From a biological point of view, the thesis characterized the differentiation and molecular profile of T_{helper} cells that play a role in inflammatory diseases, and identified cytokines and transcription factors that might be particularly important players in these processes.

From a methodological point of view, the thesis demonstrates that combining multiple approaches, data types and information allows one to gain insights into biological phenomena in a way that would otherwise not be possible.

5 ACKNOWLEDGEMENTS

There are several people who contributed to this thesis one way or another, scientifically or as friends or both. Therefore, I would like to express my gratitude to all of them, and below some of them in particular.

Jesper Tegnér: Thank you for the freedom to explore and learn what I was interested in and giving me the time to do so (even if it did not directly yield 'useful' results by the next day), for always giving the 'green light' to anything I asked, for providing an interdisciplinary environment, for being a true example of someone who understands and values both 'worlds' (biology & computation), and for trusting that I will 'figure it out' and some day defend my thesis ☺

John Andersson: Thank you for always being so positive and supportive, even when research prospects did not look promising (to say the least... ☺), and for making me believe that somehow everything will work out. From our discussions about experiments, life and whatsoever, I have always been fascinated by your good examples showing that 'clarity and simplicity' can lead to the solution in the neatest way. I hope this is an organic part of my way of thinking by now.

Angelika Schmidt: Thank you for a lot of things ☺ Thank you for your endless support, for your never-ending enthusiasm about research, for your overwhelming passion about proper experimental design & use of controls, for all I could learn from you about so many things including scientific writing & immunology, and for constantly "pressurizing" me to make sure that I am writing these lines now & that I complete the PhD ☺

Venkateshan Kannan: Thank you for giving me an insight into the 'mind of a physicist', and for making it a natural part of my thinking to always make sure that literally every little step is correct when analyzing data.

Edvard Smith: Thank you for being supportive and for contributing to a good start with my research life during my initial time in Sweden, for being inspiring to become a researcher, and for all I could learn from you in terms of doing research and thinking about research.

Thanks to all past & present members of the Unit of Computational Medicine for the lots of fika and friendly atmosphere, in particular: to **Pernilla** for always finding a quick solution and for taking care of everything, to **Rubin** for his always-good spirit and for making life at

work more fun, to **Gilad** for discussions on transcription factors but also for the great falafel, to **David** for bioinformatics discussions, to **Matilda** and **Peri** for helping with sequencing related experiments and for managing the lab making sure that it runs smoothly.

Thanks to all my friends who made my time in Sweden an extremely fun journey, in particular: to all the Jägargatan fellows including **Julian, Aga, Thibaud, Shahul, Jolinde, Pim, Frank, Oliver** for the memorable times; to **Sophie, Vilma, Michael, Aileen, Wolfgang, KT, JP** for afterworks, parties, trips and all other fun; to **Shane, Thibault, René** for the ski trip and all other fun times; to **Richard, Elliott, Becky, Maëlle** for trips, parties & all the good times; to **Matheus** for always being reliably there on ski conferences; to **Johan** for being a truly inspiring example of how to be a cool professor ☺

Thanks to my friends **Marci, Dani and Toto** for all the festivals and for the infinite list of funny things.

Thanks to **Barni** for making me interested about Sweden.

Thanks to **my grandparents** for their love & kindness.

Thanks to **Viri, Fred, Madlen, Maxim** and **Roland** for being my amazing extended family from my sister's side.

It's impossible to emphasize it too much how thankful I am to **my parents** for always supporting me and for the character building that shaped me to the person I am now.

Lehetetlen eleget hangsúlyozni mennyire hálás vagyok **szüleimnek** a szüntelen támogatásukért, és amiért sok munkával jellemet faragtak belőlem ☺

6 REFERENCES

Aarts-Riemens, T., Emmelot, M.E., Verdonck, L.F., and Mutis, T. (2008). Forced overexpression of either of the two common human Foxp3 isoforms can induce regulatory T cells from CD4+CD25– cells. Eur. J. Immunol. *38*, 1381–1390.

Baaten, B.J.G., Tinoco, R., Chen, A.T., and Bradley, L.M. (2012). Regulation of Antigen-Experienced T Cells: Lessons from the Quintessential Memory Marker CD44. Front. Immunol. *3*, 23.

Baumjohann, D., and Ansel, K.M. (2014). MicroRNA regulation of the germinal center response. Curr. Opin. Immunol. *28*, 6–11.

Becher, B., Spath, S., and Goverman, J. (2016). Cytokine networks in neuroinflammation. Nat. Rev. Immunol. *17*, 49–59.

Benoist, C., and Mathis, D. (2012). Treg cells, life history, and diversity. Cold Spring Harb. Perspect. Biol. *4*, a007021.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cao, Y., Goods, B.A., Raddassi, K., Nepom, G.T., Kwok, W.W., Love, J.C., and Hafler, D.A. (2015). Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. Sci. Transl. Med. *7*, 287ra74.

Caza, T., and Landas, S. (2015). Functional and Phenotypic Plasticity of CD4(+) T Cell Subsets. Biomed Res. Int. *2015*, 521957.

Chroneos, Z.C., Midde, K., Sever-Chroneos, Z., and Jagannath, C. (2009). Pulmonary surfactant and tuberculosis. Tuberculosis (Edinb). *89 Suppl 1*, S10-4.

Codarri, L., Gyulveszi, G., Tosevski, V., Hesske, L., Fontana, A., Magnenat, L., Suter, T., and Becher, B. (2011). RORgammat drives production of the cytokine GM-CSF in helper T cells, which is essential for the effector phase of autoimmune neuroinflammation. Nat Immunol *12*, 560–567.

Crotty, S. (2014). T follicular helper cell differentiation, function, and roles in disease. Immunity *41*, 529–542.

Cua, D.J., Sherlock, J., Chen, Y., Murphy, C.A., Joyce, B., Seymour, B., Lucian, L., To, W., Kwan, S., Churakova, T., et al. (2003). Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. Nature *421*, 744–748.

Dendrou, C.A., Fugger, L., and Friese, M.A. (2015). Immunopathology of multiple sclerosis. Nat. Rev. Immunol. *15*, 545–558.

Denis, M., and Ghadirian, E. (1990). Granulocyte-macrophage colony-stimulating factor restricts growth of tubercle bacilli in human macrophages. Immunol. Lett. *24*, 203–206.

El-Behi, M., Ciric, B., Dai, H., Yan, Y., Cullimore, M., Safavi, F., Zhang, G.X., Dittel, B.N., and Rostami, A. (2011). The encephalitogenicity of T(H)17 cells is dependent on IL-1- and IL-23-induced production of the cytokine GM-CSF. Nat Immunol *12*, 568–575.

Ferber, I.A., Brocke, S., Taylor-Edwards, C., Ridgway, W., Dinisco, C., Steinman, L., Dalton, D., and Fathman, C.G. (1996). Mice with a disrupted IFN-gamma gene are susceptible to the induction of experimental autoimmune encephalomyelitis (EAE). J. Immunol. *156*, 5–7.

Gaublomme, J.T., Yosef, N., Lee, Y., Gertner, R.S., Yang, L. V, Wu, C., Pandolfi, P.P., Mak, T., Satija, R., Shalek, A.K., et al. (2015). Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity. Cell *163*, 1400–1412.

Geginat, J., Paroni, M., Maglie, S., Alfen, J.S., Kastirr, I., Gruarin, P., De Simone, M., Pagani, M., and Abrignani, S. (2014). Plasticity of human CD4 T cell subsets. Front. Immunol. *5*, 630.

Ghoreschi, K., Laurence, A., Yang, X.P., Tato, C.M., McGeachy, M.J., Konkel, J.E., Ramos, H.L., Wei, L., Davidson, T.S., Bouladoux, N., et al. (2010). Generation of pathogenic T(H)17 cells in the absence of TGF-beta signalling. Nature *467*, 967–971.

Goverman, J. (2009). Autoimmune T cell responses in the central nervous system. Nat. Rev. Immunol. *9*, 393–407.

Haak, S., Croxford, A.L., Kreymborg, K., Heppner, F.L., Pouly, S., Becher, B., and Waisman, A. (2009). IL-17A and IL-17F do not contribute vitally to autoimmune neuro-inflammation in mice. J. Clin. Invest. *119*, 61–69.

Hartmann, F.J., Khademi, M., Aram, J., Ammann, S., Kockum, I., Constantinescu, C., Gran, B., Piehl, F., Olsson, T., Codarri, L., et al. (2014). Multiple sclerosis-associated IL2RA polymorphism controls GM-CSF production in human TH cells. Nat. Commun. *5*, 5056.

Hong, T., Xing, J., Li, L., and Tyson, J.J. (2012). A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. BMC Syst. Biol. *6*, 66.

Janeway, C.A., and Medzhitov, R. (2002). Innate immune recognition. Annu. Rev. Immunol. *20*, 197–216.

Kara, E.E., McKenzie, D.R., Bastow, C.R., Gregor, C.E., Fenix, K.A., Ogunniyi, A.D., Paton, J.C., Mack, M., Pombal, D.R., Seillet, C., et al. (2015). CCR2 defines in vivo development and homing of IL-23-driven GM-CSF-producing Th17 cells. Nat. Commun. *6*, 8644.

Kleinewietfeld, M., Manzel, A., Titze, J., Kvakan, H., Yosef, N., Linker, R.A., Muller, D.N., and Hafler, D.A. (2013). Sodium chloride drives autoimmune disease by the induction of pathogenic TH17 cells. Nature *496*, 518–522.

Kole, R., Krainer, A.R., and Altman, S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. Nat. Rev. Drug Discov. *11*, 125–140.

Kuchroo, V.K., Collins, M., al-Sabbagh, A., Sobel, R.A., Whitters, M.J., Zamvil, S.S., Dorf, M.E., Hafler, D.A., Seidman, J.G., and Weiner, H.L. (1994). T cell receptor (TCR)

usage determines disease susceptibility in experimental autoimmune encephalomyelitis: studies with TCR V beta 8.2 transgenic mice. J. Exp. Med. *179*, 1659–1664.

Langrish, C.L., Chen, Y., Blumenschein, W.M., Mattson, J., Basham, B., Sedgwick, J.D., McClanahan, T., Kastelein, R.A., and Cua, D.J. (2005). IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. J Exp Med *201*, 233–240.

Lee, P.W., Yang, Y., Racke, M.K., and Lovett-Racke, A.E. (2015). Analysis of TGF-β1 and TGF-β3 as regulators of encephalitogenic Th17 cells: Implications for multiple sclerosis. Brain. Behav. Immun. *46*, 44–49.

Lee, Y., Awasthi, A., Yosef, N., Quintana, F.J., Xiao, S., Peters, A., Wu, C., Kleinewietfeld, M., Kunder, S., Hafler, D.A., et al. (2012). Induction and molecular signature of pathogenic TH17 cells. Nat Immunol *13*, 991–999.

Levy, D.E., and Darnell, J.E. (2002). Signalling: STATs: transcriptional control and biological impact. Nat. Rev. Mol. Cell Biol. *3*, 651–662.

Mailer, R.K.W., Falk, K., and Rötzschke, O. (2009). Absence of Leucine Zipper in the Natural FOXP3Δ2Δ7 Isoform Does Not Affect Dimerization but Abrogates Suppressive Capacity. PLoS One *4*, e6104.

Mallucci, G., Peruzzotti-Jametti, L., Bernstock, J.D., and Pluchino, S. (2015). The role of immune cells, glia and neurons in white and gray matter pathology in multiple sclerosis. Prog. Neurobiol. *127–128*, 1–22.

Massagué, J. (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol. *13*, 616–630.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

McGeachy, M.J., Chen, Y., Tato, C.M., Laurence, A., Joyce-Shaikh, B., Blumenschein, W.M., McClanahan, T.K., O'Shea, J.J., and Cua, D.J. (2009). The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17-producing effector T helper cells in vivo. Nat Immunol *10*, 314–324.

McQualter, J.L., Darwiche, R., Ewing, C., Onuki, M., Kay, T.W., Hamilton, J.A., Reid, H.H., and Bernard, C.C. (2001). Granulocyte macrophage colony-stimulating factor: a new putative therapeutic target in multiple sclerosis. J. Exp. Med. *194*, 873–882.

Morikawa, M., Derynck, R., and Miyazono, K. (2016). TGF-β and the TGF-β Family: Context-Dependent Roles in Cell and Tissue Physiology. Cold Spring Harb. Perspect. Biol. *8*, a021873.

Noster, R., Riedel, R., Mashreghi, M.F., Radbruch, H., Harms, L., Haftmann, C., Chang, H.D., Radbruch, A., and Zielinski, C.E. (2014). IL-17 and GM-CSF expression are antagonistically regulated by human T helper cells. Sci Transl Med *6*, 241ra80.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. (Stanford InfoLab).

Parham, C., Chirica, M., Timans, J., Vaisberg, E., Travis, M., Cheung, J., Pflanz, S., Zhang,

R., Singh, K.P., Vega, F., et al. (2002). A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rbeta1 and a novel cytokine receptor subunit, IL-23R. J Immunol *168*, 5699–5708.

Pillai, V., Ortega, S.B., Wang, C.K., and Karandikar, N.J. (2007). Transient regulatory T-cells: a state attained by all activated human T-cells. Clin. Immunol. *123*, 18–29.

Piper, C., Pesenacker, A.M., Bending, D., Thirugnanabalan, B., Varsani, H., Wedderburn, L.R., and Nistala, K. (2014). Brief Report: T Cell Expression of Granulocyte-Macrophage Colony-Stimulating Factor in Juvenile Arthritis Is Contingent Upon Th17 Plasticity. Arthritis Rheumatol. *66*, 1955–1960.

Ponomarev, E.D., Shriver, L.P., Maresz, K., Pedras-Vasconcelos, J., Verthelyi, D., and Dittel, B.N. (2006). GM-CSF Production by Autoreactive T Cells Is Required for the Activation of Microglial Cells and the Onset of Experimental Autoimmune Encephalomyelitis. J. Immunol. *178*, 39–48.

Ramesh, R., Kozhaya, L., McKevitt, K., Djuretic, I.M., Carlson, T.J., Quintero, M.A., McCauley, J.L., Abreu, M.T., Unutmaz, D., and Sundrud, M.S. (2014). Pro-inflammatory human Th17 cells selectively express P-glycoprotein and are refractory to glucocorticoids. J Exp Med *211*, 89–104.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rolla, S., Bardina, V., De Mercanti, S., Quaglino, P., De Palma, R., Gned, D., Brusa, D., Durelli, L., Novelli, F., and Clerico, M. (2014). Th22 cells are expanded in multiple sclerosis and are resistant to IFN-β. J. Leukoc. Biol. *96*, 1155–1164.

Romagnani, S. (2014). T cell subpopulations. Chem. Immunol. Allergy *100*, 155–164.

Rothchild, A.C., Stowell, B., Goyal, G., Nunes-Alves, C., Yang, Q., Papavinasasundaram, K., Sassetti, C.M., Dranoff, G., Chen, X., Lee, J., et al. (2017). Role of Granulocyte-Macrophage Colony-Stimulating Factor Production by T Cells during Mycobacterium tuberculosis Infection. MBio *8*, e01514-17.

Sergushichev, A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv 60012.

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

Teng, M.W.L., Bowman, E.P., McElwee, J.J., Smyth, M.J., Casanova, J.-L., Cooper, A.M., and Cua, D.J. (2015). IL-12 and IL-23 cytokines: from discovery to targeted therapies for immune-mediated inflammatory diseases. Nat. Med. *21*, 719–729.

Wang, J., Ioan-Facsinay, A., van der Voort, E.I.H., Huizinga, T.W.J., and Toes, R.E.M. (2007). Transient expression of FOXP3 in human activated nonregulatory CD4+ T cells. Eur. J. Immunol. *37*, 129–138.

Wicks, I.P., and Roberts, A.W. (2015). Targeting GM-CSF in inflammatory diseases. Nat. Rev. Rheumatol. *12*, 37–48.

Wong, M.T., Ong, D.E.H., Lim, F.S.H., Teng, K.W.W., McGovern, N., Narayanan, S., Ho, W.Q., Cerny, D., Tan, H.K.K., Anicete, R., et al. (2016). A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. Immunity *45*, 442–456.

Wu, C., Yosef, N., Thalhamer, T., Zhu, C., Xiao, S., Kishi, Y., Regev, A., and Kuchroo, V.K. (2013). Induction of pathogenic TH17 cells by inducible salt-sensing kinase SGK1. Nature *496*, 513–517.

Zielinski, C.E. (2014). Autoimmunity beyond Th17: GM-CSF producing T cells. Cell Cycle *13*, 2489–2490.