

---

**The surface atlas of human naive and activated CD4<sup>+</sup> T cells  
– characterization of early T cell activation  
on a multi-omic level**

---



Dissertation  
der Fakultät für Biologie  
der Ludwig-Maximilians-Universität München

vorgelegt von  
**Anke Gräßel, geb. Fleißner**  
aus Marktleuthen

München 2016

---

**The surface atlas of human naive and activated CD4<sup>+</sup> T cells  
– characterization of early T cell activation  
on a multi-omic level**

---



Dissertation  
der Fakultät für Biologie  
der Ludwig-Maximilians-Universität München  
zur Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von  
**Anke Gräßel, geb. Fleißner**  
aus Marktleuthen

München 2016

---

Die vorliegende Dissertation wurde im Zeitraum von Januar 2013 bis März 2016 am Zentrum für Allergie und Umwelt (ZAUM) - Technische Universität und Helmholtz Zentrum München durch Herrn Prof. Dr. Carsten B. Schmidt-Weber betreut und von Herrn Prof. Dr. Benedikt Grothe von der Fakultät für Biologie der Ludwig-Maximilians-Universität München vertreten.

Erstgutachter: Prof. Dr. Benedikt Grothe

Zweitgutachter: Prof. Dr. Angelika Böttger

Tag der Abgabe: 14.04.2016

Tag der mündlichen Prüfung: 12.07.2016

#### Erklärung

Ich, Anke Gräßel, geboren am 24.10.1985 in Marktredwitz versichere hiermit an Eides statt, dass meine Dissertation selbstständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 14.04.2016

Wesentliche Teile dieser Dissertation sind in der folgenden Publikation veröffentlicht:

Graessel, A. et al. A Combined Omics Approach to Generate the Surface Atlas of Human Naive CD4<sup>+</sup> T Cells during Early T-Cell Receptor Activation. Mol Cell Proteomics 14, 2085-2102, doi:10.1074/mcp.M114.045690 (2015).

## I. Abstract

Naive CD4<sup>+</sup> T cells are the precursor cells of all effector T helper cell subsets and they form the basis of the immunologic memory. These cells provide one of the earliest cellular targets to modulate T cell activation and differentiation during the development of CD4<sup>+</sup> T cell driven immune pathologies such as autoimmune diseases and allergies, which are an increasing problem for the worlds' societies. Easy accessible cell surface proteins are responsible for the recognition of and response to signals of other cells or changes in the environment, therefore, they can be described as interesting targets for immune modulation strategies such as immunotherapy and vaccination. The aim of this dissertation is to characterize the proteomic cell surface composition of human naive CD4<sup>+</sup> T cells and their changes during T cell activation on a multi-omic level to deepen the current knowledge about these important immune cells and to identify new immune targets for the development of novel immune modulation strategies. Human naive CD4<sup>+</sup> T cells were isolated and activated with anti-CD3/anti-CD28 in a time course experiment to mimic T cell receptor engagement. The samples were analyzed via a non-targeted proteomic technique (PAL-qLC-MS/MS), a targeted flow cytometry screen and a genome-wide microarray expression analysis coupled to bioinformatics analyses. All obtained results were combined in the surface atlas of human naive and activated CD4<sup>+</sup> T cells. Out of the analyzed multi-omic datasets, the transmembrane protein c16orf54 was chosen for further investigations and tools such as monoclonal antibodies, stable expression systems and murine model organisms were generated. 229 cell surface proteins were identified and quantified on human naive and activated CD4<sup>+</sup> T cells by the proteomic techniques and 927 cell surface protein coding transcripts were detected by the transcriptomic analyses. 51 of the cell surface proteins are annotated as targets for approved drugs and further interesting cell surface targets such as solute carrier transport proteins and proteins, which were not described in the context of T cell biology before, like the transmembrane protein c16orf54, were identified by analyses of the multi-omic datasets. Newly generated investigational tools revealed that c16orf54 is not only expressed on naive and activated CD4<sup>+</sup> T cells within the compartment of immune cells in the blood.

The generated surface atlas of human naive and activated CD4<sup>+</sup> T cells can be seen as multi-omic reference guide for CD4<sup>+</sup> T cell activation, increasing the current knowledge of CD4<sup>+</sup> T cell biology. In addition, it provides a rich source of interesting immune targets, which can be investigated in the context of novel therapeutic strategies aiming to modulate reactions of the immune system during the development of CD4<sup>+</sup> T cell driven diseases.

## II. Zusammenfassung

Naive  $CD4^+$  T-Zellen sind Vorläuferzellen für alle Effektor-T Helferzell-Subtypen und bilden die Basis für das immunologische Gedächtnis. Diese Zellen stellen die frühestmöglich beeinflussbare zelluläre Zielstruktur dar, um die T-Zell Aktivierung und Differenzierung während der pathologischen Entwicklung von  $CD4^+$  T-Zell-bedingten Krankheiten wie Autoimmunkrankheiten und Allergien zu modulieren. Diese Krankheiten sind ein zunehmendes Problem für die Gesellschaft auf der ganzen Welt. Einfach zugängliche Proteine auf der Zelloberfläche sind dafür zuständig, Signale von anderen Zellen oder des sich verändernden Umfeldes zu erkennen und aufzunehmen. Aus diesem Grund sind Oberflächenproteine interessante Zielstrukturen für immunmodulatorische Strategien wie Immuntherapie und Impfung. Ziel der vorliegenden Dissertation ist die Charakterisierung der Proteinzusammensetzung auf der Oberfläche von humanen naiven  $CD4^+$  T-Zellen und deren Veränderung während der T-Zell Aktivierung basierend auf mehrstufigen Omics-Techniken. Dies soll den derzeitigen Wissenstand über diese wichtigen Immunzellen vertiefen und neue Zielstrukturen identifizieren die für die Entwicklung von innovativen immunmodulatorischen Strategien dienen können. Humane naive  $CD4^+$  T-Zellen wurden isoliert und für verschiedene Zeitspannen mit anti-CD3/anti-CD28 zur T-Zell-Rezeptor Aktivierung stimuliert. Die Proben wurden mittels einer umfassenden proteomischen Technik (PAL-qLC-MS/MS), einer gezielten durchflusszytometrischen- und einer genomweiten Microarray Expressionanalyse, gekoppelt an Bioinformatik, untersucht. Alle Ergebnisse wurden zum Oberflächenatlas für die naive und aktivierte  $CD4^+$  T-Zelle zusammengefasst. Aus den mehrstufigen Omics-Datensätzen wurde das Transmembranprotein c16orf54 für weitere Untersuchungen ausgewählt und es wurden monoklonale Antikörper, stabile Expressionssysteme und murine Modellorganismen als Hilfsmittel zur weiteren Analyse des Proteins hergestellt. Durch die proteomischen Analysen konnten 229 Oberflächenproteine auf naiven und aktivierten  $CD4^+$  T-Zellen identifiziert und quantifiziert werden und durch Transkriptomanalyse wurden 927 Transkripte beschrieben, die für Oberflächenproteine kodieren. 51 dieser Oberflächenproteine sind annotiert als Zielstrukturen für zugelassene Wirkstoffe, aber auch weitere interessante Zielstrukturen wie Solute Carrier Transportproteine und Proteine ohne bisherig beschriebenen Zusammenhang mit T-Zellen, wie das Transmembranprotein c16orf54, konnten identifiziert werden. Durch Verwendung der neu hergestellten Hilfsmittel zur Analyse des Transmembranproteins c16orf54 konnte gezeigt werden, dass dieses Protein nicht nur auf naiven und aktivierten  $CD4^+$  T-Zellen innerhalb der Immunzellpopulation im Blut exprimiert wird. Der generierte Oberflächenatlas für humane naive und aktivierte  $CD4^+$  T-Zellen, basierend auf mehrstufigen

Omics-Datensätzen, kann als Nachschlagewerk für die Aktivierung von CD4<sup>+</sup> T-Zellen angesehen werden und erweitert den derzeitigen Wissenstand zur Biologie von CD4<sup>+</sup> T-Zellen. Zusätzlich beinhaltet der Atlas eine ergiebige Liste von interessanten Zielstrukturen, die im Kontext von neuen therapeutischen Strategien zur Modulation der Immunreaktion während der Entwicklung von CD4<sup>+</sup> T-Zell-bedingten Krankheiten untersucht werden können.

## Table of contents

I. Abstract .....	4
II. Zusammenfassung .....	5
Table of contents .....	7
III. Introduction.....	12
1. Naive CD4 <sup>+</sup> T cells – the basis of the immunologic memory.....	13
1.1 Development and maturation of human T lymphocytes .....	13
1.2 Activation and differentiation of human naive CD4 <sup>+</sup> T cells.....	14
1.3 The concept of T cell plasticity, termination of T cell response and formation of the immunologic memory .....	18
2. Immune pathologies arising from CD4 <sup>+</sup> T cell failures and related immune-modulating treatment strategies .....	20
2.1 The concept of defective tolerance in autoimmune diseases and allergies.....	20
2.2 Immunotherapy – a concept to modulate immune responses.....	21
3. The surface proteome of a cell – communication bridge and grateful pool of target structures for the development of therapeutic strategies.....	23
IV. Aim of the study .....	27
V. Methods.....	28
1. Human blood donors and cell isolation.....	28
1.1 Ethical statement and study subjects.....	28
1.2 Isolation of human naive CD4 <sup>+</sup> T cells and T cell activation.....	28
2. Generation of surface atlas of human naive and activated CD4 <sup>+</sup> T cells.....	29
2.1 PAL-qLC-MS/MS .....	29
2.1.1 Cell surface protein labeling .....	30
2.1.2 Enzymatic protein digestion and glycopeptide enrichment .....	31

---

2.1.3 Liquid chromatography - tandem mass spectrometry (LC-MS/MS) .....	32
2.1.4 Database-search and label-free relative quantification of peptides .....	33
2.1.5 Data processing and identification of proteins.....	33
2.1.6 Technical validation of PAL-qLC-MS/MS.....	34
<b>2.2 Protocols for cell staining and flow cytometry analysis .....</b>	<b>34</b>
2.2.1 Detailed settings for individual experiments using flow cytometry .....	35
2.2.2 Flow cytometry-based cell surface screening and data analysis .....	36
<b>2.3 Analysis of proteomic results: unsupervised clustering by GProx and Gene ontology (GO) enrichment analysis and DrugBank target search .....</b>	<b>37</b>
<b>2.4 Transcriptomic analysis of naive and activated CD4<sup>+</sup> T cells .....</b>	<b>37</b>
2.4.1 RNA isolation and RNA quality measurement .....	37
2.4.2 Whole genome microarray and analysis of resulting transcriptomic data.....	38
2.4.3 Sequence mapping, <i>in silico</i> identification of transcripts coding for cell surface proteins and GO term analysis .....	38
<b>2.5 Combination and correlation of proteomic and transcriptomic datasets.....</b>	<b>39</b>
<b>2.6 Identification and targeted validation of cell surface proteins which were not mentioned in the context of T cell biology before .....</b>	<b>40</b>
2.6.1 Validation of candidates by qPCR.....	40
2.6.2 Validation of candidates by Western blot .....	41
<b>3. Analysis of transmembrane protein c16orf54, a cell surface protein which was not described in the context of T cell biology before .....</b>	<b>42</b>
<b>3.1 Generation, production and testing of rat and mouse monoclonal antibodies against human and murine c16orf54 .....</b>	<b>42</b>
3.1.1 Peptide-immunization of rats and mice and hybridoma generation .....	42
3.1.2 Testing of hybridoma supernatants against biotinylated peptides via ELISA.....	43
3.1.3 Isolation of murine naive CD4 <sup>+</sup> T cells .....	43
3.1.4 Suitability test of hybridoma supernatants (flow cytometry and Western blot) .....	44
3.1.5 Direct labeling of antibody and antibody-peptide competition assay .....	45
<b>3.2 Generation of expression systems for recombinant expression of c16orf54.....</b>	<b>46</b>
3.2.1 Cloning of different overexpression vectors .....	46

---

3.2.1.1 Amplification of inserts .....	46
3.2.1.2 Ligation and Transformation .....	47
3.2.1.3 Colony PCR and isolation of plasmids.....	47
3.2.2 Expression of c16orf54 in mammalian cells .....	48
3.2.3 Expression and purification of soluble forms of c16orf54 in insect cells.....	48
3.2.4 Immunoprecipitation of recombinant c16orf54.....	49
<b>3.3 Generation of a CRISPR/CAS mediated knockout mouse.....</b>	<b>50</b>
3.3.1 Cloning of vectors containing different sgRNAs .....	50
3.3.2 <i>In vitro</i> activity test of the different sgRNAs and T7 endonuclease assay.....	51
3.3.3 Generation of T7-PCR templates for <i>in vitro</i> transcription of sgRNAs.....	52
3.3.4 <i>In vitro</i> transcription and purification of Cas9 and sgRNAs .....	52
3.3.5 Embryo microinjection, culture and transfer .....	53
<b>3.4 Genotyping of mouse lines .....</b>	<b>53</b>
<b>VI. Results.....</b>	<b>55</b>
<b>1. Generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells.....</b>	<b>55</b>
<b>1.1 Establishment, validation and technical monitoring of PAL-qLC-MS/MS technique..</b>	<b>56</b>
1.1.1 Influence of oxidation and biotinylation process .....	57
1.1.2 Validation of protein expression via flow cytometry in parallel to PAL-qLC-MS/MS sample preparation .....	59
1.1.3 Assessment of donor variability by comparing the protein expression patterns.....	59
<b>1.2 PAL-qLC-MS/MS-based cell surface glycoproteome of human naive and activated CD4<sup>+</sup> T cells .....</b>	<b>61</b>
1.2.1 Strategic evaluation of raw qLC-MS/MS results and implementation of Trypsin and PNGase F fractions into the cell surface glycoproteome .....	61
1.2.2 Analysis of cell surface protein expression patterns by unsupervised clustering and Gene Ontology enrichment analysis of resulting expression clusters .....	66
1.2.3 Differences and similarities between naive and activated CD4 <sup>+</sup> T cells .....	68
1.2.4 Identification and validation of cell surface proteins, which were not co-cited with T cell biology before .....	69

---

<b>1.3 Validation and extension of the cell surface glycoproteome via a targeted flow cytometry-based cell surface screen.....</b>	<b>71</b>
1.3.1 Cluster analysis of protein expression patterns detected via flow cytometry .....	72
1.3.2 Comparison of surface glycoproteome and results of the flow cytometry screen and implementation of flow cytometry screening results into proteomic cell surface atlas..	73
<b>1.4 <i>In silico</i> selection and examination of cell surface protein-coding transcripts based on the whole genome microarray analysis .....</b>	<b>74</b>
<b>1.5 Combination and comparison of proteomic and transcriptomic results .....</b>	<b>77</b>
1.5.1 The surface atlas of human naive and activated CD4 <sup>+</sup> T cells.....	78
1.5.2 Differential expression analysis of the detected targets from the combined data results of transcriptome and proteome analysis .....	80
1.5.3 Correlation of PAL-qLC-MS/MS and whole genome microarray analysis.....	81
<b>1.6 The distribution of members of the SLC (solute carrier transporters) protein family on naive and activated CD4<sup>+</sup> T cells .....</b>	<b>82</b>
<b>2. c16orf54 – a cell surface protein which was not described in the context of T cell biology before.....</b>	<b>87</b>
<b>2.1 Suitability screening of anti-mouse- and anti-human c16orf54 antibodies (hybridoma supernatants) for flow cytometry.....</b>	<b>87</b>
<b>2.2 Assessment of anti-human c16orf54 antibody specificity and distribution of c16orf54 on human PBMCs.....</b>	<b>90</b>
<b>2.3 Suitability screening of anti-human c16orf54 antibody (hybridoma supernatants) for Western blot.....</b>	<b>93</b>
<b>2.4 Strategy and proof of the <i>in vitro</i> functionality of sgRNAs for CRISPR/CAS-mediated knockout mice.....</b>	<b>96</b>
<b>VII. Discussion.....</b>	<b>98</b>
<b>1. Generation of a cell surface atlas of human naive and activated CD4<sup>+</sup> T cells – technical advantages and limitations .....</b>	<b>98</b>
<b>2. Potential of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells for future studies .....</b>	<b>105</b>
<b>3. c16orf54 – a novel described cell surface protein on cells of the immune system ....</b>	<b>109</b>
<b>References.....</b>	<b>114</b>

---

<b>Materials .....</b>	<b>124</b>
<b>1. Reagents, media, buffers, enzymes, cell lines and kits .....</b>	<b>124</b>
<b>2. Antibodies.....</b>	<b>128</b>
<b>3. Primer and oligos .....</b>	<b>129</b>
<b>4. Buffers and stocks (self-made).....</b>	<b>132</b>
<b>5. Consumable material .....</b>	<b>134</b>
<b>6. Instruments.....</b>	<b>135</b>
<b>7. Software and databases .....</b>	<b>137</b>
<b>Table of Figures .....</b>	<b>138</b>
<b>List of Tables.....</b>	<b>140</b>
<b>Abbreviations .....</b>	<b>141</b>
<b>Acknowledgement.....</b>	<b>145</b>
<b>Publications .....</b>	<b>147</b>
<b>Poster &amp; Oral Presentations.....</b>	<b>148</b>
<b>Scholarships, Grants and Awards .....</b>	<b>150</b>
<b>Appendix .....</b>	<b>151</b>
<b>Supplement .....</b>	<b>154</b>

### III. Introduction

Organisms need a fast-reacting and specific protection system against invading pathogens. Nevertheless, it is also very important that this protection system is well controlled and does not overreact in safe and harmless situations. This well-balanced task is essential for the survival of all living organisms and it is fulfilled by the immune system being composed of organs, tissues and cells - interacting within a complex network inside the body. The immune system can be divided into an innate and an adaptive part. The innate immune system is responsible for the first line of defense, reacting against infections upon the very first contact to e.g. microbes at mucosal barriers. This part of the immune system is comprised of physical (skin) and chemical (antimicrobial agents produced by epithelia) barriers, proteins circulating in the blood stream (complement system associated proteins), cells with phagocytic features like macrophages and neutrophils, antigen presenting cells (APCs) like different types of dendritic cells (DCs) and also natural killer (NK) cells.<sup>1</sup> Recently, a new type of cell was identified, which is also associated with the innate immune system. These cells were named innate lymphoid cells (ILCs) and they mirror the functions of T effector cells, but do not express a T cell receptor.<sup>2,3</sup> The reaction of the innate branch of the immune system is very fast as it reacts within the first hours upon invasion. However, it represents an unspecific defense reaction, as the structures that are recognized during infection are common structures called "pathogen-associated molecular patterns" (PAMPs). PAMPs are shared among related groups of microbes, leading to the concept that involved immune cells might not be able to discriminate between different microbes. The innate immune system is also not intended to install long-lasting immunity. However, the components of the innate immune system can induce inflammation very fast by recruiting and activating cells of adaptive immunity, underlining the dependence of the adaptive immune system on the activation of innate immunity.<sup>4</sup> The adaptive branch of the immune system fulfills the task of a more specific defense combined with the aim to create the important long-lasting memory type immunity, which should protect the organism when a second contact to an immune invader takes place. Two different responses of the adaptive immune system, humoral and cell-mediated immunity, are the key components of this so-called second line of defense, which gets activated later compared to the innate immune system. Surface structures of extracellular microbes can be targeted and effectively eliminated by antibodies (humoral immunity). The cell-mediated part of the adaptive immune system is guided by lymphocytes like B and T cells. These cells are capable of specifically recognizing processed antigens from invaders presented on the surface of APCs with their specialized B- and T cell receptors (BCR; TCR)

leading to subsequent activation by receptor engagement. Another important characteristic of B and T cells is the high diversity of their specific membrane receptors BCR or TCR, creating the possibility to discriminate between different determinants or epitopes of many antigens. This, by definition called, lymphocyte repertoire is created by the variability of antigen-binding sites of the TCR and BCR originating from somatic gene recombination. It gives the immune system of a single individual the chance to detect and distinguish  $10^7 - 10^9$  different antigenic determinants and fight against a large number of immune invaders. But the negative side of this great variability is the risk that every novel created receptor might have the potential to recognize the body's own "self-antigens" or harmless environmental determinants like allergens. This would lead to unnecessary and dangerous activation of the immune system directed against the autologous system or harmless environmental structures. Therefore, different mechanisms of tolerance exist, which should protect the body from such conditions. This balance between fast reaction against harmful invaders and no reaction or tolerance induction against harmless structures needs to be maintained permanently.<sup>5</sup> A cell type, which is intensively participating in maintaining this important balance and creating the immunologic memory, is the naive  $CD4^+$  T cell (T helper cell) and especially its T helper cell subsets (Th1, Th2, Th9, Th17, Th22, Tfh, Treg), which arise by the activation and differentiation of the naive  $CD4^+$  T cell.<sup>6,7</sup>

## **1. Naive $CD4^+$ T cells – the basis of the immunologic memory**

### **1.1 Development and maturation of human T lymphocytes**

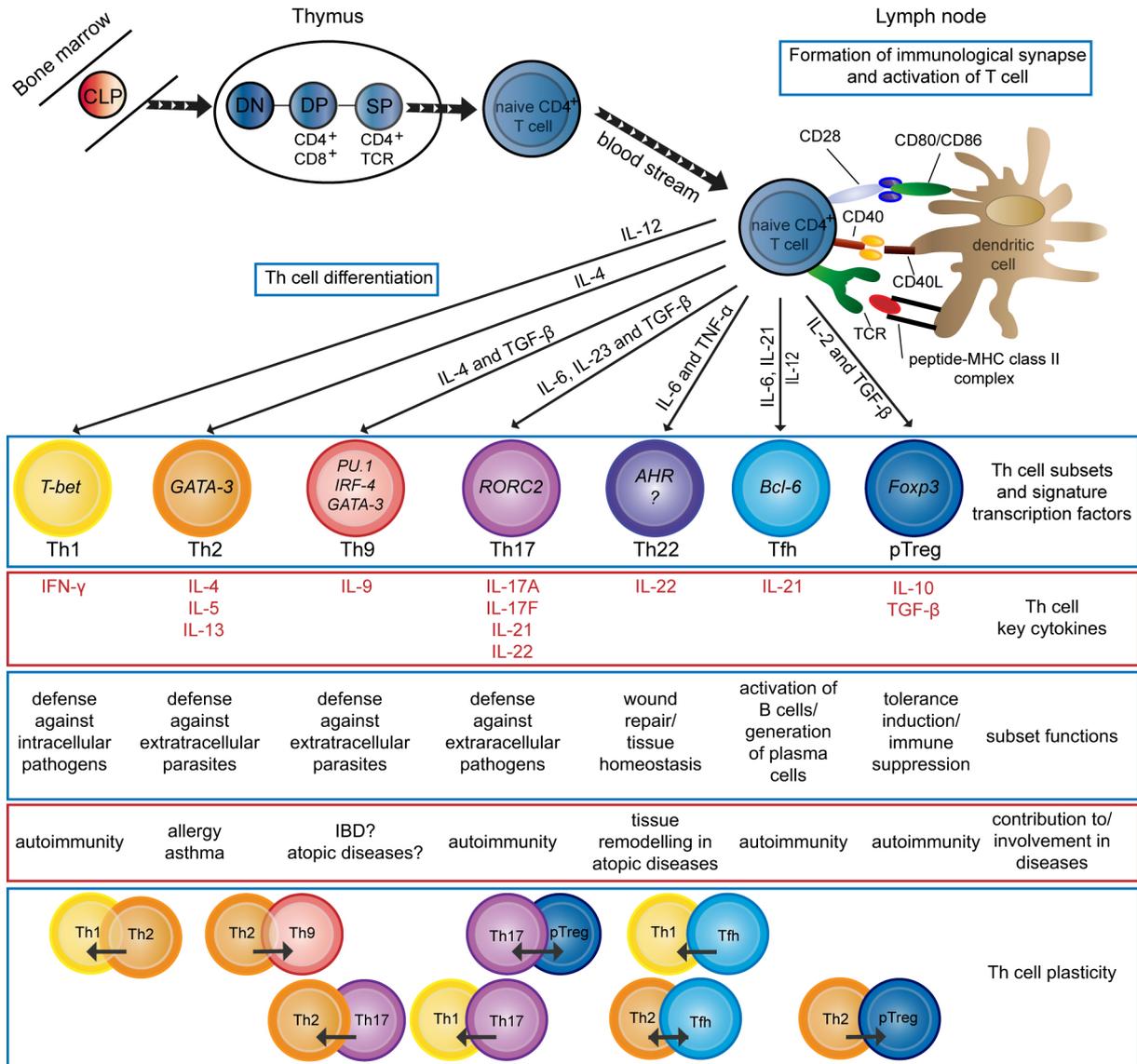
T cell development and maturation are very well characterized processes, which were described extensively before.<sup>8,9</sup> In brief, naive  $CD4^+$  T cells arise from hematopoietic stem cells originating in the bone marrow. These stem cells develop into common myeloid progenitors or common lymphoid progenitors (CLP) in the bone marrow. Later on, the myeloid lineage gives rise to erythrocytes, platelets, basophils, eosinophils, monocytes, DCs and neutrophils, whereas the lymphoid progenitors can develop into B cells, T cells and NK cells (and some types of DCs). The lymphoid progenitors stay either in the bone marrow to further mature into the B cell lineage or they migrate to the thymus as T cell precursors. The lineage commitment inside the bone marrow is guided by cell surface receptors, the accessibility of distinct genomic loci and involved transcription factors. In case of the T cell lineage, it is the interplay between Notch-1 and the transcription factor GATA-3, amongst others, which obliges the lymphoid progenitor to become a T cell precursor.<sup>10</sup> Interleukin (IL)-7 is also a very important key player at this stage, which is needed by the T cell precursors for prolifera-

tion in the bone marrow and the thymus (Fig. 1). These precursor cells enter the thymus as immature so-called thymocytes, harboring the TCR genes in their germline composition. Starting from this double-negative (DN) thymocyte state (no expression of CD3, CD4, CD8, TCR,  $\zeta$ -chain) as pro-T cell, these cells will form either a  $\gamma\delta$ - or an  $\alpha\beta$ -TCR by TCR gene rearrangement, coordinated by the proteins Rag-1 and Rag-2. The  $\alpha\beta$ -T cells will give rise to MHC class II (major histocompatibility complex) restricted CD4<sup>+</sup> T cells or MHC class I restricted CD8<sup>+</sup> T cells. During the DN state, the pro-T cells develop into pre-T cells and arrange and express their pre-TCR. The cells will then transit into the double-positive (DP) state (CD4<sup>+</sup>/CD8<sup>+</sup>) and built up a functional TCR complex comprised of a functional  $\alpha\beta$ -TCR, CD3 and  $\zeta$ -chain on the cell surface. These DP T cells get in contact with self-antigens in the thymus, which are induced by the transcription factor AIRE (autoimmune regulator) and presented on MHC molecules by e.g. epithelial cells. A selection process then takes place at this stage. DP T cells that do not recognize any “self antigen-self MHC-complex” undergo apoptosis. T cells, which bind self-antigens only with low-avidity are positively selected, survive and become single positive (SP) immature T cells. They can become either CD4<sup>+</sup> T helper cells, if they recognize the antigen presented on a MHC class II molecule, or cytotoxic CD8<sup>+</sup> T cells, if they recognize the antigen presented on a MHC class I molecule. DP T cells, which recognize and strongly bind self-antigens, are dangerous for the autologous system. They can trigger autoimmune diseases and therefore these cells are negatively selected, meaning that they undergo apoptosis.<sup>8</sup> Another described possibility is the generation of natural regulatory T cells (thymically derived regulatory T cells, tTreg) within this selection process. A potential slightly higher avidity to self-antigens than the one described for positive selection might lead to the development of tTregs, which retain central tolerance. The positively selected and surviving SP T cells leave the thymus as mature naive T cells. They still have not encountered a foreign antigen, meaning that they are seen as antigen-inexperienced, and they could be present in the circulation or could be found in peripheral lymphoid organs. Mature naive T cells can survive for 1-3 months, but they need survival signals to stay viable. These survival signals originate from low avidity interactions with presented self-antigens and Interleukin-7 (IL-7).<sup>11</sup>

## 1.2 Activation and differentiation of human naive CD4<sup>+</sup> T cells

Naive CD4<sup>+</sup> T cells recirculate throughout the body via the blood stream to lymphoid organs. When naive CD4<sup>+</sup> T cells enter lymphoid organs like the lymph nodes they get in contact with APCs presenting ingested and processed antigens, potential recognizable by the TCR of the naive CD4<sup>+</sup> T cell. A suitable antigen loaded on a MHC class II molecule on the APC and the following TCR engagement with the T cell is the first signal, which is needed by the naive

CD4<sup>+</sup> T cell for its activation. By close interaction, both cells form the immunologic synapse that is essential to properly exchange communication signals (Fig. 1). CD3 and the ζ-chain transduce this signal from the TCR to start the process of T cell activation.<sup>12</sup>



**Figure 1: Development and differentiation of the naive CD4<sup>+</sup> T cell and plasticity of T helper cell subsets.** Common lymphoid progenitors (CLP) migrate from the bone marrow to the thymus to give rise to DN T cell precursors (DN= double negative, CD4<sup>-</sup>/CD8<sup>-</sup>). These cells mature to DP (double-positive, CD4<sup>+</sup>/CD8<sup>+</sup>) and then SP (single positive, either CD4<sup>+</sup> or CD8<sup>+</sup>) immature T cells and leave the thymus as mature naive T cells. Naive CD4<sup>+</sup> T cells can get activated by contact to antigen presenting cells (e.g. dendritic cells) and via additional cytokines from the microenvironment, they can differentiate into different T helper (Th) cell subsets in the lymph node. These T helper cell subsets are characterized by the expression of signature transcription factors (italic letters) and secretion of key cytokines, which equip them with distinct functional properties within the immune defense mechanism. But the different T helper cell subsets are also involved in or even contribute to pathologic conditions. A certain plasticity between some of the T helper cell subsets was described, which is characterized by a shift from one subset to another or the formation of a hybrid cell which is able to express transcription factors and/or cytokines of different T helper cell subsets (arrows indicate the direction of a possible shift). (Tfh=T follicular helper cell, pTreg=peripheral derived regulatory T cell, IBD=inflammatory bowel disease).

Co-stimulatory molecules, expressed by the APCs, are additionally necessary for the activation. Cell surface proteins of the B7-family (CD80, CD86) are expressed on professional APCs and engage with CD28 expressed on the T cell and deliver the important co-stimulatory signal (second signal). Activated T cells then express the CD40 ligand (CD40L) on their surface and bind to CD40, which is mainly expressed on APCs. This contact enhances the expression of the B7 co-stimulatory molecules on the APC even more and amplifies the T cell activation (Fig. 1). These reactions on the surface of the T cell also initiate the start of the intracellular signaling cascade leading e.g. to the activation of the PI3-kinase, the Akt kinase and also the Ras/ERK MAP kinase pathway. Anti-apoptotic proteins are upregulated and therefore ensure T cell survival and proliferation. Other co-stimulatory molecules can be CD2<sup>13-15</sup> and different integrins, as well as different receptors of the tumor necrosis factor receptor (TNFR) superfamily for example. The activated CD4<sup>+</sup> T cells then undergo changes in their surface protein profile and start the secretion of cytokines, which is followed by their proliferation and differentiation into effector and memory CD4<sup>+</sup> T cells. The induction of surface molecules like CD69, CD25 (IL-2 Receptor  $\alpha$ ) and CD40L as early activation markers, which are expressed within hours up to one day, are already well described.<sup>16,17</sup> The secretion of IL-2 also starts very early, within the first hours after T cell receptor engagement. IL-2 fulfills autocrine as well as paracrine functions, by promoting the survival, proliferation and differentiation of the activated T cell. This process is also called clonal expansion, which means that single antigen specific T cells highly proliferate and increase their frequency from 1 specific naive T cell in  $10^5 - 10^6$  lymphocytes to 1 specific activated T cell in 100 – 1000 lymphocytes. These T cells are now developing into T effector cells to eliminate the antigen and help the body to clear the infection.<sup>18,19</sup>

These effector T cells can now be divided into different T helper cell subsets with specialized abilities and functions: Th1, Th2, Th9, Th17, Th22, Tfh (T follicular helper cells), pTreg (peripheral derived regulatory T cells).<sup>1</sup> The T helper cell subsets can be defined by their transcription factor profile and the release of specific cytokines. An overview about the differentiation possibilities of a human naive CD4<sup>+</sup> T cell is given in Fig. 1. As stated before, the T cell needs different signals to get activated (antigen recognition and T cell receptor engagement, co-stimulatory molecule interaction). To differentiate into a T helper cell subset, a third signal is needed, which is generated by the microenvironment (local derived factors from tissue, cytokines which are released by the engaged APC). Th1 and Th2 cells were the first described Th cell subsets.<sup>20</sup> Th1 differentiation relies on IL-12 as trigger, which is derived by the DC, but this process could also be pushed by type I interferons as well as the cytokines IL-18, IL-23 and IL-27.<sup>21</sup> T-bet is the signature transcription factor of Th1 cells<sup>22</sup> and it was demonstrated to be essential for the IFN- $\gamma$  production as Th1 key cytokine<sup>23,24</sup>, to fight against infections emerged by the invasion of intracellular living pathogens<sup>25</sup>, such as

*Mycobacterium tuberculosis*. Th1 cells are able to enhance the anti-microbial actions of macrophages to clear infections. The differentiation into Th2 cells depends on the co-activation with IL-4, which then leads to signal transduction by STAT6 and induction of GATA-3 expression, the signature transcription factor of Th2 cells.<sup>26</sup> This subset is induced in the body to fight against parasitic infections with extracellular living parasites<sup>27,28</sup> and mediates its anti-parasitic actions via the release of IL-4, IL-5 and IL-13, the Th2 key cytokines.<sup>29</sup> Th2 cells are capable of stimulating IgE production, recruiting and activating eosinophils and mast cells to attack the parasites. A combination of the cytokines IL-6, IL-23, and TGF- $\beta$  leads to the differentiation of naive CD4<sup>+</sup> T cells into Th17 cells<sup>30-33</sup>, which are characterized by the expression of RORC2 and STAT3, leading to the secretion of IL-17, IL-22 and IL-21.<sup>34-37</sup> Th17 cells are key players in infections, which are induced by extracellular bacteria and fungi.<sup>38-41</sup> These cells can recruit neutrophils and also monocytes, which in turn take up the microbes for elimination and are strong inducers of epithelial defense mechanisms.

The next Th cell subset was initially described as IL-9 producing Th2 cell, but this concept was refined and it was proven that the main source of IL-9 producing cells was a new subset, which was then named Th9 cell.<sup>42-44</sup> The cytokines IL-4 and TGF- $\beta$  promote the differentiation into Th9 cells and the described signature transcription factors of this subset are GATA-3, IRF-4 and PU.1. The current assumption of the Th9 cell function is, as it is for Th2 cells, the defense against extracellular living parasites.<sup>45</sup>

The Th22 cell subset is one of the latest described Th cell subsets<sup>46-48</sup> and this subset relies on IL-6 and TNF $\alpha$  for differentiation. These cells are characterized by the expression of IL-22, but not IL-17 distinguishing them from IL-22 producing Th17 cells. At the moment a signature transcription factor for these cells cannot be named, it was only shown that the aryl hydrocarbon receptor (AHR) seems to be important for expression of IL-22.<sup>47</sup> Functionally, Th22 cells are important for wound repair and maintenance of tissue homeostasis.<sup>48,49</sup>

Follicular T helper (Tfh) cells are another important subset, located in the follicles of lymph nodes.<sup>50</sup> They can originate from naive CD4<sup>+</sup> T cells by stimulation with IL-6, IL-21 and IL-12, but also from already activated T cell subsets that seem to be not fully committed. The differentiation into this subset is guided by the contact with activated germinal center B cells and the engagement of ICOS on the T cell with ICOS ligand on the B cell. Tfh cells secrete IL-21 as their key cytokine, which is required for the development of the germinal center in the lymph node<sup>51-53</sup> and the generation of plasma cells.<sup>54-56</sup> This cell subset is also influencing the isotype switching reaction in B cells and their signature transcription factor is Bcl-6.<sup>57-59</sup>

Besides the T helper cell subsets with a clear mission to protect against external harm, regulatory T (Treg) cells are part of the CD4<sup>+</sup> T cell family, but with the mission to preserve the balance of the immune system between inflammation and tolerance and to prevent autoimmunity.<sup>60-62</sup> Tregs engage different mechanisms to induce tolerance. They secrete IL-10

and TGF- $\beta$  as inhibitory cytokines<sup>63</sup> or inhibit the activating-capacity of APCs by the direct interaction of CTLA-4 on the Treg and a B7-molecule on the APC.<sup>64</sup> Furthermore, they induce apoptosis in self-reactive effector T cells and by their massive demand on IL-2, they simply deprive effector T cells of IL-2, which they also need for proliferation and clonal expansion.<sup>65</sup> Tregs can be divided into two groups, the thymically derived Tregs (see section 1.1), which express CD4, CD25 and Foxp3 as their signature transcription factor<sup>66,67</sup> and the peripheral derived regulatory T cells (pTreg)<sup>68-70</sup>, which differentiate from naive CD4<sup>+</sup> T cells after contact to an antigen in the periphery.<sup>71,72</sup> pTregs can additionally be subdivided into three groups, one is the group of T regulatory type 1 (Tr1) cells, characterized by expression of CD4 and CD25 but not Foxp3 and secretion of high levels of IL-10.<sup>73,74</sup> The second group is called Th3 (expression of CD4, CD25 and Foxp3 and secretion of high TGF- $\beta$  levels)<sup>75,76</sup> and the third group (CD4<sup>+</sup>/CD25<sup>+</sup>/Foxp3<sup>+</sup>) is also characterized by the expression of CD4, CD25 and Foxp3 but secretes IL-10 as well as TGF- $\beta$ .<sup>77,78</sup> To induce regulatory T cells, the cytokines IL-10 (Tr1)<sup>73,74</sup>, TGF- $\beta$  (Th3)<sup>75,76</sup> and TGF- $\beta$  in combination with IL-2 and retinoic acid are described.<sup>79-82</sup>

These different T helper cell subsets fulfill a variety of tasks due to their different capabilities. Although they are very specialized to act in different infectious situations and support the immune system to clear infections, they show a high level of plasticity.<sup>83</sup>

### **1.3 The concept of T cell plasticity, termination of T cell response and formation of the immunologic memory**

Th cell subsets are defined groups of cells attributed with special transcription factor and cytokine profiles as well as specialized effector tasks within the immune system. But it was repeatedly shown that Th cell subsets are able to change their phenotype to adapt to changes within their microenvironment.<sup>83</sup> This adaptive concept of T cell lineage flexibility is called T cell plasticity (Fig. 1) and is either achieved by conversion of one Th cell subset into a different Th cell subset or forming a kind of “hybrid-cell”, which is capable of expressing transcription factors and cytokines of two lineages at the same time. For Th17 cells it was shown that prolonged culture induces IFN- $\gamma$  production<sup>84</sup>, the key cytokine of Th1 cells. During inflammatory as well as normal conditions, a hybrid T cell population, which is producing IFN- $\gamma$  as well as IL-17, can be found.<sup>85-89</sup> Under the control of IL-4, Th17 cells can also acquire a kind of Th2 phenotype.<sup>90,91</sup> The Th17/Treg axis is also well described as being highly plastic in both directions, able to shift between an inflammatory and a regulatory state, pTregs were for example shown to be able to change towards a Th17-like phenotype under inflammatory conditions in the presence of IL-2 and IL-1- $\beta$ <sup>92,93</sup> or IL-6 and TGF- $\beta$ .<sup>94</sup> Conversely, a murine study proved that Th17 cells can TGF- $\beta$ -dependent adopt a Tr1 pheno-

type with functional capacity to favor the clearance of an inflammation.<sup>95,96</sup> Regarding the Th1/Th2 axis, studies tended to describe these as more stable subset phenotypes, because the signature transcription factors of these subsets are suppressing each other<sup>97</sup>, but it was also demonstrated that during viral infections, stable Th2 cells are able to produce IFN- $\gamma$  in addition to IL-4.<sup>98</sup> Th2 cells are likewise also able to produce IL-9 under the influence of TGF- $\beta$ <sup>43</sup>, might produce IL-10 like Tr1 cells<sup>99</sup> and by IL-21 stimulation they can adapt a Tfh phenotype.<sup>100</sup> Also in case of Tfh cells, it is possible to speculate about a very plastic phenotype, as these cells were shown to be able to express IL-4 during helminth infection<sup>100-102</sup>, but also IFN- $\gamma$  during bacterial infection of the Th1 type.<sup>102</sup> This plasticity concept of different Th cell subsets, which are shifting towards another, is more likely than definite terminally differentiated Th cell lineages. This might be a mechanism of the immune system to quickly adapt to changing situations of infectious and inflammatory conditions in a very effective way and it would be a mechanism to compensate the decrease of possible de novo responses due to thymus involution in adulthood.<sup>83</sup>

A mechanism, which also needs to be very effective, is the generation of a long-term immunologic memory. Effector T cells of all subsets only survive until the antigen is cleared and no more survival signals like IL-2 and anti-apoptotic proteins are present anymore. The reaction is shut down and >90 % of antigen-specific T cells, which originated by clonal expansion, will die by apoptosis. This decline is very important for the homeostasis of the immune system when the antigen is eliminated.<sup>1</sup> But the immune system is built up in a way, that upon a second contact to a pathogen, it reacts faster and more effectively to protect the body and this is also the basis for vaccination strategies. Memory CD4<sup>+</sup> T cells are generally characterized by the expression of CD45RO and high expression of CD127 (IL-7R) and CD44. They are a heterogeneous group of different cell subsets, composed of central memory T cells (T<sub>cm</sub>), which are able to circulate through the blood and enter secondary lymphoid organs, effector memory T cells (T<sub>em</sub>)<sup>103</sup> and tissue resident memory T cells (T<sub>rm</sub>).<sup>104,105</sup> The development of these cells is still not completely understood and different concepts about this process are discussed in the field, whereas more is known for the CD8<sup>+</sup> T cell population than for the CD4<sup>+</sup> T cell compartment. One theory is that memory T cells develop from effector memory T cells, which are not dying after the elimination of the stimulating antigen.<sup>106,107</sup> Another concept is that effector and memory T cells evolve in parallel to each other from the naive CD4<sup>+</sup> T cell upon activation by asymmetric cell division.<sup>108</sup> The important features of memory T cells, which enables them to survive for months or even years, is high expression of anti-apoptotic proteins, low proliferation due to slow cycling and recurring IL-7 consumption.<sup>109,110</sup>

## 2. Immune pathologies arising from CD4<sup>+</sup> T cell failures and related immune-modulating treatment strategies

All T helper cell subsets have defined roles within the defense mechanisms of the body against immune invaders. If the T cell development and differentiation is running under perfect conditions, the immune system homeostasis is well balanced by clearance of antigens from dangerous pathogens, installation of long-lasting immunologic memory and tolerance against self-antigens and harmless molecules such as allergens. But if errors occur during T cell development or the T helper cell subsets fail to fulfill their designated tasks, T cells contribute to the formation of different immune pathologies (Fig. 1).

### 2.1 The concept of defective tolerance in autoimmune diseases and allergies

Autoimmune diseases are characterized by a deficit in the establishment or maintenance of self-tolerance. This deficit leads to a homeostatic imbalance of the immune system between control and activation, which could be either systemic or organ-specific, depending on the distribution of the recognized self-antigen, followed by tissue injury. As described before (section 1.1), during the maturation of T cells in the thymus, T cells are tested for the recognition of self-antigens to identify and eliminate those cells that strongly bind to self-antigens. If the transcription factor AIRE, which induces the expression of self-antigens within the thymus, is not expressed or mutated and not functional, T cells with specificity for self-antigens escape the negative selection process and enter the circulation, where they systemically cause the autoimmune polyendocrine syndrome (APS). Patients affected by this disease show lymphocyte-mediated injury of several endocrine organs<sup>111,112</sup> caused by a deficiency in the establishment of central T cell tolerance. Peripheral tolerance is established at sites of peripheral tissue by different possible mechanism. As described before, T cells need more than one signal to get activated. If the co-stimulation is missing or CTLA-4 instead of CD28 on the T cell is engaging with B7-molecules (CD80/CD86) on the APC, this could induce anergy (functional unresponsiveness) or apoptosis in T cells which bind to presented self-antigens in the periphery.<sup>64</sup> Another mechanism of peripheral tolerance is the capability of Tregs to suppress the action of other T cells, in this case self-reactive T cells, e.g. by secretion of suppressing cytokines such as TGF- $\beta$  and IL-10.<sup>63</sup> Autoimmunity might therefore develop because of deficits during the thymic T cell selection process, apoptosis defects in self-reactive T cells, malfunction of inhibitory receptors, low numbers of Tregs or functional defects in the Treg compartment. Typical autoreactive T cell mediated (mostly Th1 and Th17 cells) autoimmune diseases are e.g. rheumatoid arthritis, multiple sclerosis, Type 1 diabetes mellitus and inflammatory bowel disease. Genetic predisposition and environmental factors

such as infections also contribute to the formation of autoimmune diseases and in all cases the cause of such diseases could not be attributed to a single factor.

Allergies also display a disease, which is characterized by a type of defective tolerance of the immune system. Allergic diseases can be manifested as allergic rhinitis (hay fever) or asthma in the airway compartment, eczematous reactions in skin and food allergies in the gastrointestinal tract. Allergens, which are common, *per se* harmless environmental antigens such as proteins from house dust mite, food, pollen and animal dander or chemicals like the drug penicillin, trigger an immune response in allergic patients, although these allergens are not harmful for the body and should normally be tolerated. Naive CD4<sup>+</sup> T cells get activated by allergen-presenting APCs and differentiate into Th2 cells. By secretion of IL-4 and IL-13, the Th2 cells induce isotype switching to IgE in allergen-specific B cells. IL-5 secreted by Th2 cells furthermore recruits and activates eosinophils and IL-13 stimulates excessive mucus production by epithelial cells.<sup>113</sup> IgE then binds to its receptor on mast cells and basophils leading to cell activation when two specific IgEs at the same time are crosslinked by the allergen. Upon activation, mast cells secrete lipid mediators, histamine and pro-inflammatory cytokines and therefore cause vasodilation, vascular leakage, bronchoconstriction, inflammation and tissue damage in the end – the symptoms of allergic reactions. The underlying pathogenesis of allergy is represented by a complex interplay of genetic predisposition, deviations in the immune system and various environmental factors.

## **2.2 Immunotherapy – a concept to modulate immune responses**

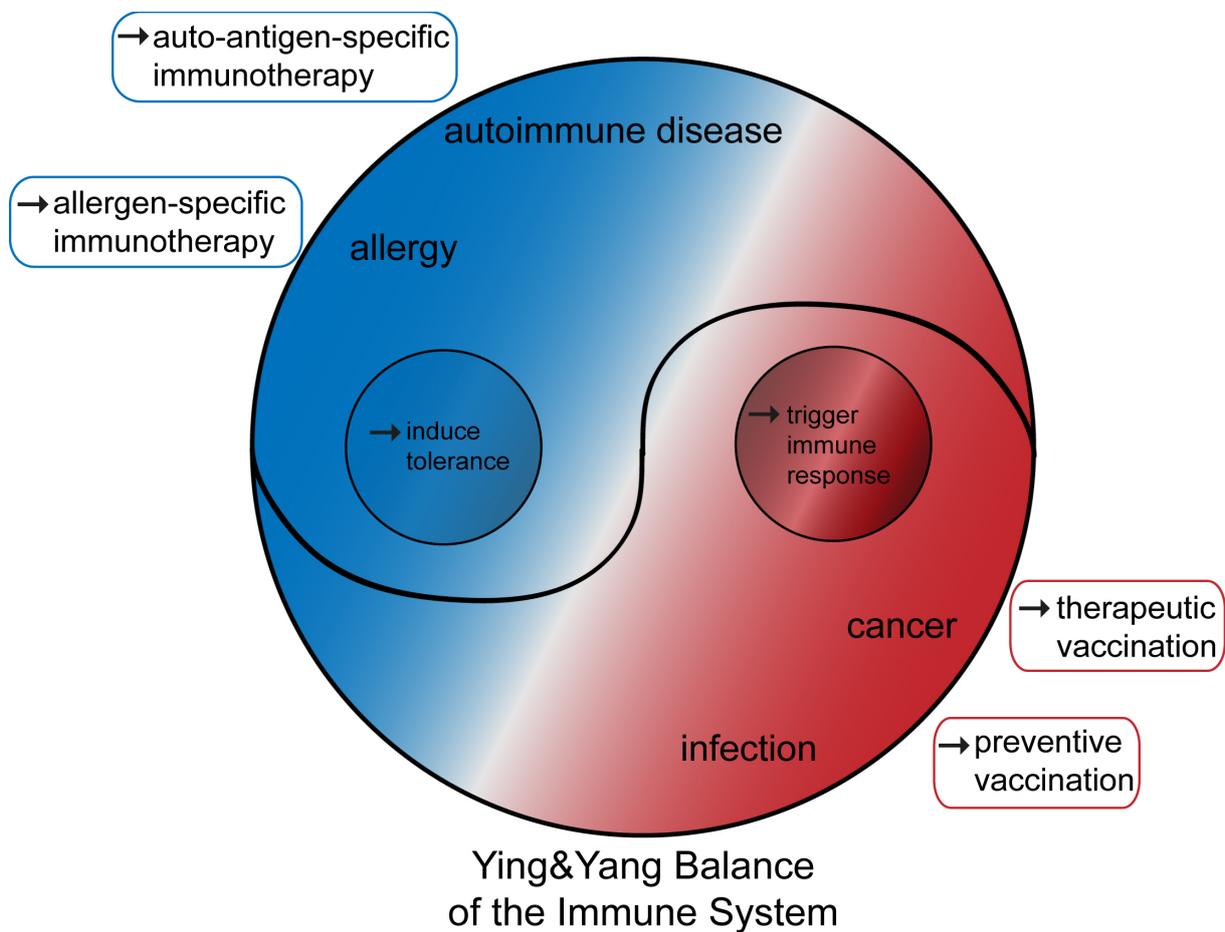
There are two main concepts behind the term immunotherapy, meaning that this type of therapy should either trigger the immune system to respond against a certain structure or induce tolerance of the immune system for a certain structure (Fig. 2).

Vaccination, which was already discovered and applied in the 18<sup>th</sup> century by Edward Jenner, is an effective strategy to trigger a desired immune response. The idea behind preventive and therapeutic vaccination is different, but both should result in the initiation of an immune response.

The effectiveness of preventive vaccination was shown by the eradication of several infectious diseases. Vaccines are composed of living attenuated pathogens, inactivated (killed) pathogens (bacteria or virus) or purified antigens from pathogens in combination with adjuvants or immunomodulators to activate the T cell response of antigen-specific CD4<sup>+</sup> T cells. The most effective vaccines are capable of inducing long-lasting immunologic memory in the

T and B cell compartment to be able to react very fast upon contact with the vaccine-targeted pathogen to clear the antigen as effectively as possible.<sup>114,115</sup>

Therapeutic vaccination for cancer treatment is a challenging strategy to trigger immune responses against a tumor. There are many existing anti-cancer vaccination strategies, based on different vaccines and delivery methods, which are sometimes combined with adjuvants. Vaccines could for example be peptides or full-length proteins (identified on the tumor cells) as epitopes for CD8<sup>+</sup> cytotoxic T cells, autologous APCs loaded with tumor antigens or DNA of autologous killed tumor cells. Viruses and autologous APCs are considered to be carriers for the vaccine.<sup>116,117</sup> At the moment there are only two FDA approved cancer vaccines (prostate cancer, melanoma) available in the USA, leaving room to improve vaccination strategies for cancer therapy (Sipuleucel-T, talimogene laherparepvec (T-VEC), <http://www.cancer.gov>).



**Figure 2: Balance of the immune system.** Several diseases like allergies, autoimmune diseases, infections and cancer disturb the balance of the immune system by changing the ratio between tolerance and activation. To reconstitute the balance within the immune system, therapeutic strategies aim to either induce tolerance in the case of allergy and autoimmune diseases via allergen-/auto-antigen-specific immunotherapy or trigger an immune response in the case of cancer and infections via therapeutic or preventive vaccination strategies.

Tolerance inducing immunotherapy against specific auto-antigens or allergens is the second immunotherapy concept, which could be applied as a therapeutic strategy to treat or even heal autoimmune diseases and allergies. Regarding the treatment of autoimmunity, strategies to alter the identified pathologic cytokine patterns of the different diseases are currently the focus. But in diabetes mellitus type 1 for example, an innovative research strategy to induce tolerance with agonistic insulin mimetopes is currently being investigated.<sup>118</sup> Allergen-specific immunotherapy (AIT) has already been applied for a long period of time and is based on desensitization of allergic patients by repeated administration of low doses of allergens subcutaneously or orally/sublingual.<sup>119-121</sup> The mechanisms of AIT to improve clinical symptoms are based on the reduction of allergen-specific IgE, the increase of allergen-specific IgG4 and the induction of allergen specific regulatory T and B cells. If these points are achieved via the therapy, the number of Th2 cells and ILC2s, together with their effector cytokines IL-4, IL-5 and IL-13, decreases and IgE antibody production is lowered leading to a consecutive reduction of mast cell, basophil and eosinophil activation or at least activation to a lower extent.<sup>122-129</sup> Some patients benefit from this kind of immunotherapy, but the tolerance induction phase takes a long time and many patients stop the therapy for this reason. Unfortunately, the efficacy of the therapy for each individual patient can still not be predicted during the treatment.<sup>130</sup> However, the search for efficacy predicting biomarkers is ongoing and improving.<sup>131</sup> These predictive biomarkers are actually urgently needed as a motivation for the patient to complete the therapy as well as in aiding the physician's decision about discontinuing therapy and possibly searching for alternative treatments.

Allergic and autoimmune diseases are a rising problem all over the world, but especially in westernized countries. The European Academy for Allergy and Clinical Immunology (EAACI) states that 150 million people are already affected in Europe and this number is still increasing. A recent analysis of data of the last 30 years also showed that incidence and prevalence for autoimmune diseases are increasing.<sup>132</sup> New effective therapies need to be generated to not only treat symptoms by immunosuppression, but to specifically induce long-lasting tolerance in affected patients and restore the balance of the immune system.

### **3. The surface proteome of a cell – communication bridge and grateful pool of target structures for the development of therapeutic strategies**

“ome” such as in proteome, transcriptome or genome are datasets derived from so-called “omics” techniques, which aim for a description, as complete as possible, of its central molecules of interest like proteins, transcripts or genes. High-throughput methods such as mass spectrometry, genome-wide microarray expression analysis, RNA-sequencing, or whole

genome sequencing enable the generation of enormous amounts of data points and are inevitable techniques nowadays. Studies focusing on T cell differentiation already investigated the transcriptome and epigenome of different T helper cell subsets earlier and revealed new differentiation key drivers for these molecular levels.<sup>133-135</sup> While the immunologic as well as proteomic community repeatedly requested proteomic approaches investigating components of the immune system<sup>136,137</sup>, especially T cell immunology, two drafts of the human proteome were published in 2014<sup>138,139</sup>, which demonstrated and emphasized on the necessity of large-scale proteomic data. Many proteomic studies interested in T cell biology were conducted in the Jurkat T cell line.<sup>140-142</sup> A first proteomic study on primary human T cells was published in 2001 and contained 91 proteins, which were identified by metabolic labeling, 2D gel electrophoresis and MALDI-TOF mass spectrometry.<sup>143</sup> Changes of the nuclear proteome of activated human cord blood CD4<sup>+</sup> T cells upon IL-4 stimulation<sup>144</sup>, as well as changes of the phosphoproteome of human primary CD4<sup>+</sup> T cells, which were stimulated with anti-CD3 for 5 min<sup>145</sup>, were also investigated before. These studies were conducted with whole cell lysates, nuclear extracts or focused on the analysis of the phosphorylation status of the proteins. Another very interesting cellular compartment to analyze is the cell surface, as the communication of a cell with its surrounding environment takes place here. The cell surface proteome is the unity of cell surface proteins. This unity is qualified to recognize, take up, process and translate signals from the environment of a cell into intracellular signaling events and to further generate departing signals, which leave the cell through the cell surface, back towards the environment. Cell surface proteome focusing approaches, such as the cell surface capture technology<sup>146</sup>, enabled the description of the surface proteome of stem cells<sup>147</sup>, mesenchymal stromal cells<sup>148</sup>, murine adipocytes in obesity<sup>149</sup> and B cell lines originating from lymphomas.<sup>150</sup> Recently, a descriptive compilation of cell surface proteome datasets of 78 human and murine cell types, which can be browsed in an open-access interactive database, was also published.<sup>151</sup>

It was shown before, that such non-targeted cell surface focusing techniques do not only present detailed descriptions of cell surface compositions, but are capable of identifying proteins, which could be valuable targets for the development of diagnostic markers or new therapeutic approaches.<sup>152</sup> An important advantage of cell surface proteins compared to other intracellular structures is their easy accessibility. Drugs, targeting cell surface molecules do not need to be able to pass the cell membrane to initiate their designated changes and a variety of possibilities for targeted therapies, depending on the type of cell surface protein, are given. Recombinant produced cytokines can be administered to target cytokine receptors of T helper cell subsets and induce specific actions.<sup>153,154</sup> Antibodies can target cytokine receptors to block their accessibility for endogenously produced cytokines (Table 1) or target cytokines to block their interaction with receptors. Also the class of small

molecules, which can either serve as agonist or antagonist, is a valuable class of drugs, capable of targeting cell surface proteins. An overview about currently approved therapeutic antibodies and therapeutic antibodies in clinical trials targeting cell surface molecules of abnormal behaving CD4<sup>+</sup> T helper cells or their products is presented in Table 1.

Targeted structure/cell	Mechanism of action	Type of molecule	Brand/Generic name	Approval date/Clinical Phase	Clinical application
TCR	anti-CD3	mouse IgG2a mAb	OKT3 (muro-nomab-CD3)	approved 1986 (withdrawn by sponsor 2012)	transplant rejection
all T cells	anti-T cell sera	rabbit, anti-thymoglobulin	Thymoglobulin (anti-thymocyte globulin)	1998	transplant rejection
thymocytes	anti-IL2R	chimeric IgG1	Simulect (basiliximab)	1998	transplant rejection
CD4	anti-CD4 mAb	humanized IgG4 mAb	ibalizumab	phase III	HIV infection
CD26	anti-CD26 mAb	mouse IgG2b mAb	Begecina (be-gelomab)	regulatory review (EU)	Graft-versus-host-disease
CD28	CTLA-Fc co-stimulatory blockade	receptor fusion	Orenica (abatacept)	2005	RA
CCR5	anti-CCR5 mAb	humanized IgG4 mAb	PRO-140, PA14	phase II/III	HIV infection
Th2 cells	anti-IL-4R $\alpha$ mAb	human IgG4 mAb	dupilumab	phase III	atopic dermatitis, asthma
Th1/Th17 cells	anti-p40 (IL-12/IL-23) mAb	human IgG1 mAb	Stelara (ustekinumab)	2009	psoriasis, Crohn's disease
Th1/Th17 cells	anti-TNF	humanized IgG1 Fab, pegylated	Cimzia (certolizumab pegol)	2009	Crohn's disease
Th1/Th17 cells	anti-TNF mAb	human IgG1 mAb	Simponi (golimumab)	2009	psoriatic arthritis, AS
Th1/Th17 cells	anti-TNF mAb	human IgG1 mAb	Humira (adalimumab)	2003	RA
Th1/Th17 cells	anti-TNF mAb	chimeric IgG1	Remicade (infliximab)	1999	Crohn's disease
CNS homing lymphocytes	anti- $\alpha$ 4 $\beta$ 1/ $\beta$ 7 integrin receptor	humanized IgG4 mAb	Tysabri (natalizumab)	2006	ms
gut homing lymphocytes	anti- $\alpha$ 4 $\beta$ 7/ $\alpha$ E $\beta$ 7 integrin receptor	humanized IgG1 mAb	Entyvio (vedolizumab)	2014	ulcerative colitis, Crohn's disease
gut homing lymphocytes	anti- $\alpha$ 4 $\beta$ 7/ $\alpha$ E $\beta$ 7 integrin receptor	humanized IgG1 mAb	etrolizumab	phase III	ulcerative colitis, Crohn's disease
Th2 cell product	anti-IL-5 mAb	humanized IgG1 mAb	Nucala (mepolizumab)	2015	eosinophilic asthma
Th2 cell product	anti-IL-5 mAb	humanized IgG4 mAb	reslizumab	regulatory review (EU, USA)	asthma
Th2 cell product	anti-IL-13 mAb	humanized IgG4 mAb	lebrikizumab	phase III	asthma

Targeted structure/cell	Mechanism of action	Type of molecule	Brand/Generic name	Approval date/Clinical Phase	Clinical application
Th2 cell product	anti-IL-13 mAb	human IgG4 mAb	tralokinumab	phase III	asthma
Th17 cell differentiation	anti- IL-6 R mAb	humanized IgG1	RoActemra (tocilizumab)	2009	RA
Th17 cell differentiation	anti-IL-23 p19 subunit	human IgG1 mAb	guselkumab	phase III	psoriasis
Th17 cell differentiation	anti-IL-23 p19 subunit	humanized IgG1 mAb	tildrakizumab	phase III	psoriasis
Th17 cell differentiation	anti- IL-6 mAb	human IgG1 mAb	sirukumab	phase III	RA, giant cell arteritis
Th17 cell differentiation	anti- IL-6 R mAb	humanized IgG2 mAb	SA237	phase III	NMO and NMO spectrum disorders
Th17 cell product	anti-IL-17A/A mAb	human IgG1 mAb	Cosentyx (secukinumab)	2015 phase III phase II	psoriasis, AS, PA, RA RRms AD
Th17 cell product	anti-IL-17RA mAb	human IgG2 mAb	brodalumab	regulatory review (EU)	psoriasis
Th17 cell product	anti-IL-17A/A; A/F mAb	humanized IgG4 mAb	ixekizumab	regulatory review (EU)	psoriasis

**Table 1: List of therapeutic antibodies currently approved/under regulatory review or in phase II/III clinical trials against cell surface structures on T cells and their secreted products (excluded are antibodies for cancer therapy).** The list provides the target structure of the drug, the type of acting therapeutic molecule, the brand name (if available) and the name of the active agent, the year of approval (either US or EU) or phase of clinical trial and the clinical application (status February 2016). (AS=ankylosing spondylitis, Fab=fragment antigen binding, mAb=monoclonal antibody, ms=multiple sclerosis, NMO=neuromyelitis optica, PA=psoriatic arthritis, RA=rheumatoid arthritis, RRms=relapsing, remitting multiple sclerosis, AD=atopic dermatitis) ([www.fda.gov/](http://www.fda.gov/); [www.ema.europa.eu/ema](http://www.ema.europa.eu/ema;); <sup>155</sup>)

The naive CD4<sup>+</sup> T cell is the precursor of all effector T helper cell subsets and forms the basis for the immunologic memory, thus providing one of the earliest cells to target and modulate during the development of CD4<sup>+</sup> T cell driven pathologies. As there are still many unmet needs in the field of CD4<sup>+</sup> T cell driven diseases such as autoimmune diseases and allergies and room for improvement of vaccination strategies e.g. in the cancer therapy field, the generation of a cell surface atlas of human naive and activated CD4<sup>+</sup> T cells would create a basis for the development of new therapeutic and vaccination strategies. In addition, the surface proteome of naive and activated CD4<sup>+</sup> T cells would give a detailed overview about proteins located on the cell surface of these important cells of the immune system and describe how the expression of these proteins changes during early T cell activation, thus enlarging the current understanding of CD4<sup>+</sup> T cell biology.

## IV. Aim of the study

T cell driven diseases such as allergies, asthma and autoimmune diseases represent an increasing problem for society. Not only the personal burden for the patients such as dealing with symptoms and pain or coping with a reduced quality of life is a permanent problem, but also the costs for the health care systems are a growing problem for social systems. Therefore, it is a key issue to understand the differences of T cell biology between healthy individuals and the abnormal state in pathologic manifestations. However, to gain such a deep knowledge regarding T cell pathologies and to be able to develop effective therapeutic strategies, we first need to better understand how a healthy T cell network is maintained.

This dissertation focuses on the T cell surface as cellular compartment of interest, because the cell surface is the line of communication between a cell and its environment. It is responsible for the exchange of signals by recognizing and responding to intracellular and extracellular changes. Furthermore, the cell surface proteins, which are located in or associated with the plasma membrane, are not only receiving signals, they are also responsible and capable themselves of initiating further signaling events. Cell surface proteins are mostly easy accessible and different strategies, such as blocking, mimicking and activating, to target them from outside the system already exist. Therefore, cell surface proteins are valuable structures regarding the issue of target discovery for drug development.

Thus, the aim of this dissertation is to describe the cell surface protein composition of freshly isolated human naive CD4<sup>+</sup>/CD45RA<sup>+</sup> T cells from healthy blood donors and to describe in detail how this composition changes during T cell activation. It will combine new and already well-established quantitative omics technologies coupled to straightforward bioinformatics analyses, leading to the generation of a cell surface atlas of human naive and activated CD4<sup>+</sup> T cells.

The results of this dissertation will assist in gaining a deeper insight into changes happening during T cell activation on the surface of naive CD4<sup>+</sup> T cells in a healthy *ex vivo* system. On the one hand side it will provide a huge number of cell surface proteins, which are known to be expressed on T cells. Due to the implementation of non-targeted approaches on the other hand side, it might also have the potential to identify cell surface proteins, which were not related to T cell biology before. All results obtained on transcriptomic as well as proteomic level will be combined in a cell surface atlas and this cell surface atlas of naive and activated CD4<sup>+</sup> T cells will present a surface reference guide for T cell activation.

## V. Methods

### 1. Human blood donors and cell isolation

#### 1.1 Ethical statement and study subjects

Voluntary human subjects, taking part in this study (local ethics committee of the Technical University Munich under ethic board approval number 2877/10), gave their written informed consent to donate peripheral blood. Their total serum immunoglobulin E (IgE) levels were measured (Dermatology Department, Klinikum rechts der Isar, München) and the levels of specific IgE against the common allergens *Dermatophagoides pteronyssinus*, cat danders, wheat flour, celeriac, timothy grass, secale cereal, birch, hazel, and mugwort were checked by radio-allergen-sorbent-test (RAST). Study subjects were declared as non-atopic when total IgE levels were lower than 50 kU/L, results of the RAST were negative and subjects declared that they do not have a history of atopic diseases.

#### 1.2 Isolation of human naive CD4<sup>+</sup> T cells and T cell activation

Peripheral blood mononuclear cells (PBMCs) were freshly isolated from heparinized peripheral blood of human healthy, non-atopic donors by density gradient centrifugation. The heparinized blood was diluted 1:1 with D-PBS and 25 mL of this dilution were layered on 10 mL of Lymphoprep. This gradient was centrifuged without brake at 975 x g for 15min. The PBMCs were collected after centrifugation by aspirating the distinct PBMC band at the interface between Lymphoprep and plasma. The mononuclear cells were washed with D-PBS containing 5 mM EDTA at least four times, by repeated resuspension and centrifugation steps (step 1: 10min at 515 x g, step 2-4: 10min at 290 x g). A magnetic activated cell sorting (MACS)-based method was used to isolate naive CD4<sup>+</sup> T cells (Naive CD4<sup>+</sup> T cell isolation Kit II human), which are characterized by the expression of CD45RA and no expression of CD45RO. In order to calculate the cell number, an aliquot of PBMCs was diluted in a Trypanblue solution, counted with a Neubauer cell counting chamber and Trypanblue positive cells were excluded from the total cell number. The cells were resuspended in autoMACS running buffer (20 µL MACS buffer for 1 x 10<sup>7</sup> PBMCs), biotinylated antibody cocktail was added (5 µL antibody cocktail for 1 x 10<sup>7</sup> PBMCs) and incubated for 15min on ice. The magnetic anti-Biotin beads were added (15 µL anti-Biotin beads for 1 x 10<sup>7</sup> PBMCs) to the PBMC-antibody mix and incubated for 10min on ice to bind the antibody-labeled cells.

Naive CD4<sup>+</sup> T cells were isolated by negative selection through an automated magnetic column (autoMACS, program: DEPLETES). To guarantee a maximum purity of the naive CD4<sup>+</sup> T cell population, another purification with a second isolation kit was carried out. Hereby, the isolated cells from the first step were counted, resuspended in autoMACS running buffer (40 µL MACS buffer for 1 x 10<sup>7</sup> T cells) and incubated with magnetic beads coupled to anti-CD45RO antibody (CD45RO microbeads) (10 µL beads for 1 x 10<sup>7</sup> T cells) for 20min on ice. The cells were washed with D-PBS containing 2 mM EDTA, centrifuged for 10min at 290 x g and resuspended in autoMACS running buffer (500 µL MACS buffer for 10<sup>8</sup> cells). Negative magnetic separation was carried out with the autoMACS (program: DEPLETES). The purity of the untouched naive CD4<sup>+</sup> T cell population was analyzed by flow cytometry and the cells were only used if a purity of at least 95 % was reached (section 2.2).

For experiments regarding the activation of naive CD4<sup>+</sup> T cells, the wells of a 24-well plate were coated with 0.75 µg anti-CD3 in 1 mL D-PBS for 3h at 37 °C. Immediately after aspiration of the coating solution, 1 x 10<sup>6</sup> naive CD4<sup>+</sup> T cells were seeded in 1.5 mL of AIMV containing 0.75 µg/mL soluble anti-CD28 per well. The cells were incubated at 37 °C and 5 % CO<sub>2</sub> in a humidified incubator.

Regarding the whole genome microarray analysis, 1 x 10<sup>6</sup> naive CD4<sup>+</sup> T cells were activated with the T cell activation/expansion Kit according to the manufacturers' recommendations with a bead-to-cell ratio of 1:2 in 1 mL of AIMV medium for 3h.

## 2. Generation of surface atlas of human naive and activated CD4<sup>+</sup> T cells

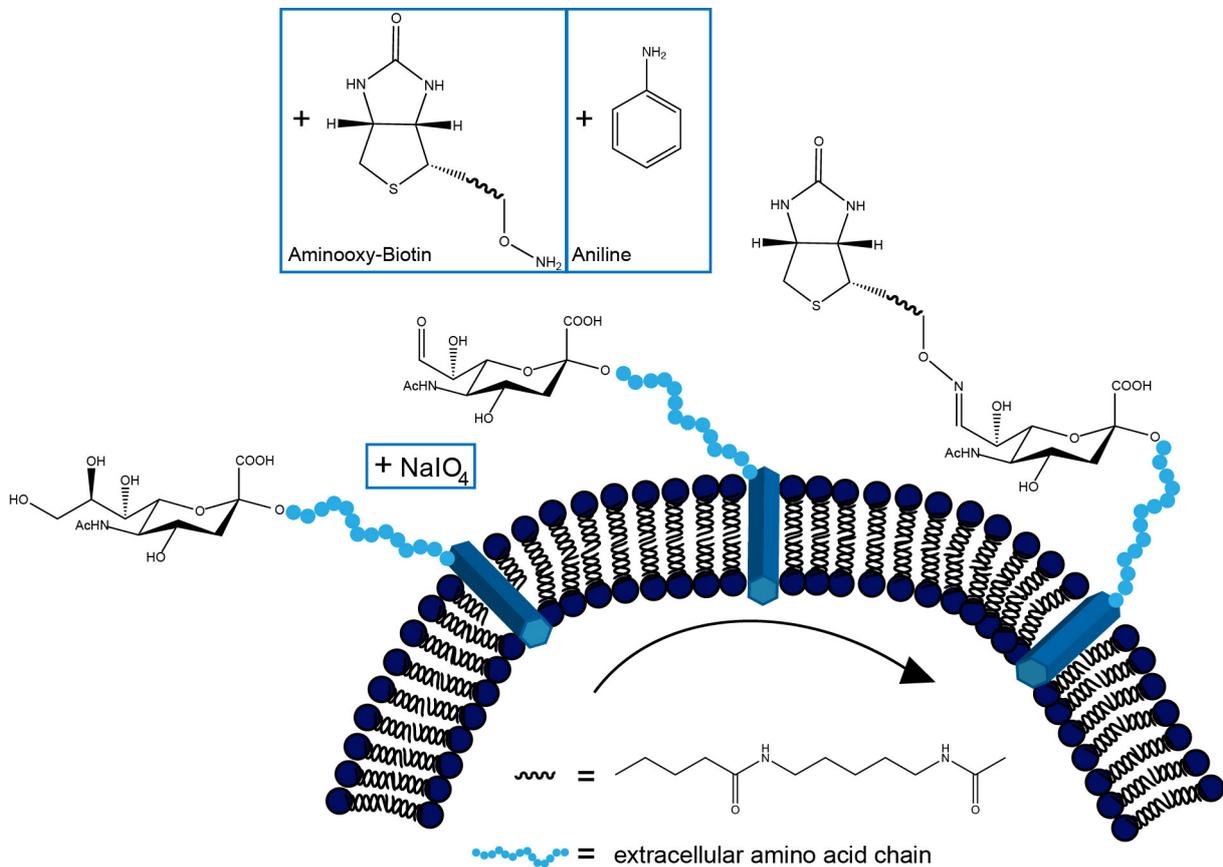
Isolated human naive CD4<sup>+</sup> T cells of 19 study subjects were used in total to generate and validate the surface atlas of human naive CD4<sup>+</sup> T cells. These cells were either taken in their naive form or activated (see 1.2) for different periods of time. The surface atlas is built on three pillars (see Fig. 4 in section: Results), containing two proteomic and one transcriptomic experimental setup coupled to *in silico* analysis. Details for each pillar are given in the following sections.

### 2.1 PAL-qLC-MS/MS

The previously described technology PAL - periodate oxidation and aniline-catalyzed oxime-ligation<sup>156</sup> (Fig. 3) was adapted for the use of primary human naive CD4<sup>+</sup> T cells and used to label and enrich cell surface proteins. To identify and quantify cell surface proteins, the PAL samples were subjected to quantitative LC-MS/MS (liquid chromatography-tandem mass spectrometry). Cellular samples of four human subjects (donor D1-D4) were processed and analyzed with the non-targeted PAL-qLC-MS/MS technique (Fig. 4, blue track). The samples

of every donor included naive CD4<sup>+</sup> T cells and naive CD4<sup>+</sup> T cells activated for 3, 6, 12, 24 and 48h in a time course experiment with anti-CD3/anti-CD28.

### 2.1.1 Cell surface protein labeling



**Figure 3: Cell surface glycoprotein labeling via PAL (periodate oxidation and aniline catalyzed oxime ligation) technique.** NaIO<sub>4</sub> is used to oxidize the sialic acid, which is a possible modification of glycosylated cell surface proteins. Aniline catalyzes the oxime ligation of Aminooxy-Biotin to the aldehyde, which was introduced by oxidation at the sugar residues. Aminooxy-Biotin forms a covalent bond and therefore a stable linkage to the glycosylated cell surface protein. This reaction is conducted as a one-pot reaction. (adapted from <sup>156</sup>)

All the following steps of this process<sup>157</sup> were carried out at 4 °C on ice.  $8 \times 10^6$  naive or activated CD4<sup>+</sup> T cells of every donor and time point ( $n=4$  donors, 6 time points of each donor, 48 samples in total) were centrifuged for 10min at 1500 x g directly after the isolation, respectively at the end of the indicated activation time. Cell pellets were washed twice with 1 mL ice-cold labeling buffer (PBS with CaCl<sub>2</sub>/MgCl<sub>2</sub>, pH 6.7) and centrifuged at 1500 x g for 10min. After washing, the cell pellets were resuspended in 1 mL of oxidation/biotinylation mix in an one-pot-reaction, consisting of 1 mM NaIO<sub>4</sub>, 100 μM Aminooxy-Biotin and 10 mM aniline in labeling buffer. The samples were incubated while rotating in the dark for 30min at 4 °C to oxidize the glycosylation and form the link to Aminooxy-Biotin. To stop the biotinylation reaction, glycerol was added to a final concentration of 1 mM and samples were

incubated on the rotator for further 5min at 4 °C. The cells were washed with 1 mL ice-cold wash buffer (PBS with CaCl<sub>2</sub>/MgCl<sub>2</sub>, pH 7.4) and biotinylation labeling efficiency was analyzed by flow cytometry (Streptavidin-PE staining). The washed cell pellets were resuspended in 250 µL lysis buffer (1 % NP-40, 10 mM NaCl, 10 mM Tris pH 7.6, 2 x EDTA-free complete protease inhibitor cocktail in ddH<sub>2</sub>O) and stored at -20 °C until the cell membrane fractions were prepared.

During the establishment of the labeling process of living primary cells, a suitable concentration of NaIO<sub>4</sub> needed to be found to guarantee a possible maximum of cell viability and to show that this treatment does not have adverse effects on the protein expression or the correct detection of the protein expression. Therefore, different concentrations of NaIO<sub>4</sub> were tested in the oxidation/biotinylation mix and the effect on cell viability and protein expression were checked via flow cytometry (section 2.2).

### 2.1.2 Enzymatic protein digestion and glycopeptide enrichment

The preparation of the cell membrane fractions for mass spectrometry analysis<sup>157</sup>, started with defrosting the lysed cells on ice (approximately 15min in total), complemented by vortexing every two minutes. The raw lysates were cleared by centrifugation at 6000 x g at 4 °C for 10min. While the pellet was discarded, the supernatant was diluted 1:5 in wash buffer. In parallel, 60 µL of Strep-Tactin beads (Superflow 50 % suspension) per sample were washed for three times with 500 µL wash buffer and centrifuged at 1000 x g for 1min in LoBind tubes. The complete sample was transferred to the LoBind tubes containing the prewashed beads and incubated on a rotator at 4 °C for 2h to enable the binding of biotinylated cell surface proteins to high-affinity Strep-Tactin beads. To pellet the protein-bead complexes, the samples were centrifuged at 1000 x g for 1min. All following incubation and washing steps were conducted in a total buffer volume of 200 µL per sample and the centrifugation steps were carried out at 2000 x g for 2min (RT). For the first washing step of the beads the wash buffer was supplemented with 0.2 % NP-40. The second step was carried out with the wash buffer containing 0.5 % SDS. Then, the beads were incubated in wash buffer (supplemented with 0.5 % SDS and 100 mM DTT) for 30min at RT. This was followed by centrifugation and then the beads were washed with UC buffer (6 M Urea, 100 mM Tris-HCl pH 8.5). The beads were incubated in UC buffer (supplemented with 50 mM iodoacetamide) at RT for 30min to start the process of alkylation of the proteins. After these incubation steps the beads were washed with the following buffers: UC buffer, 5 M NaCl, 100 mM Na<sub>2</sub>CO<sub>3</sub> pH 11.5, 50 mM Tris-HCl pH 8.5. To digest the proteins, the protein-bead complexes were resuspended in 40 µL 50 mM Tris-HCl pH 8.5 after the last centrifugation and 1 µg sequencing-grade modified trypsin was added. This was incubated in a gently shaking

Thermomixer at 37 °C overnight. The samples were then centrifuged to collect the tryptic fraction of the peptides in the supernatant, which were transferred into a new LoBind tube. The beads were resuspended in 40 µL of the 50 mM Tris-HCL pH 8.5 buffer and centrifuged. The resulting supernatant was pooled with the first tryptic fraction and the tryptic peptides were stored at -20 °C. The beads were then washed with 40 µL 1 x G7 buffer and after centrifugation they were resuspended in 20 µL of 1 x G7 buffer. To release the glycopeptides, 500 U of glycerol-free PNGase F were added to each sample and these mixtures were incubated in a gently shaking Thermomixer at 37 °C for 6h. The samples were centrifuged after the second enzymatic digestion and the supernatant, containing the glycopeptides, was transferred into a new LoBind tube. 20 µL of 1 x G7 buffer was added to the beads and the mix was resuspended. After centrifugation, the resulting supernatant was pooled with the first PNGase F fraction and stored at -20 °C until analysis by mass spectrometry. The tryptic and the PNGase F fraction were stored and measured separately.

### 2.1.3 Liquid chromatography - tandem mass spectrometry (LC-MS/MS)

The LC-MS/MS measurement<sup>157</sup> was performed in cooperation with the Research Unit Protein Science, Helmholtz Center Munich (Dr. Stefanie Hauck and Dr. Christine von Toerne). The LC-MS/MS analyses were performed as described before<sup>158</sup> on the LTQ-Orbitrap XL Mass Spectrometer with the following adjustments: To separate the peptides before the separation by reversed phase chromatography, a nano trap column (300 µm inner diameter × 5 mm, packed with Acclaim PepMap100 C18.5 µm, 100 Å) was used. The reversed phase chromatography (PepMap, 25 cm, 75 µm ID, 2µm/100 Å pore size) was operated on a RSLC (Ultimate 3000). A nonlinear 170 min LC gradient from 5 to 31 % of buffer B (98 % acetonitrile and 0.1 % formic acid) at 300 nl/min flow rate followed by a short gradient from 31 to 95 % buffer B in 5min and an equilibration for 15min to starting conditions, was chosen. The 10 most abundant peptide ions from the MS prescan were selected for fragmentation in the linear ion trap, if they were at least doubly charged and exceeded an intensity of at least 200 counts. High-resolution (60,000 full-width half maximum) MS spectra were acquired in the Orbitrap ranging from a mass of 300 to 1500 Da, during fragment analysis. One microscan was recorded with fill times in the FT (MS) set to 0.5s and in the Ion Trap (MSn) to 0.1s. The automatic Gain Control (AGC) targets were set to  $1 \times 10^6$  (MS) and  $1 \times 10^4$  (MSn), respectively. The MS proteomics data have been deposited to the ProteomeXchange Consortium<sup>159</sup> via the PRIDE partner repository with the dataset identifier PXD001432.

#### 2.1.4 Database-search and label-free relative quantification of peptides

The database-search and label-free relative quantification of peptides<sup>157</sup> was performed in cooperation with the Research Unit Protein Science, Helmholtz Center Munich (Dr. Stefanie Hauck and Dr. Christine von Toerne). The RAW files (Thermo Fisher Scientific) were analyzed using the Progenesis LC-MS software (version 4.0, Nonlinear Dynamics), as described by Hauck *et al.*<sup>160</sup> The following changes were made: The search engine Mascot (Matrix Science, Release number 2.4) was utilized to identify peptides, assuming tryptic digestion (one missed cleavage was allowed), a fragment ion mass tolerance of 0.6 Da and a parent ion tolerance of 10 ppm. Carbamidomethylation was set as fixed modification and allowed variable modifications were methionine oxidation and asparagine or glutamine deamidation. The spectra were searched against the Ensembl human database (Release 69; 100,607 sequences)<sup>161</sup> and the common contaminants keratin and albumin were excluded. The Mascot-integrated decoy database search was performed by using the Percolator algorithm. An average peptide false discovery rate of < 1 % was then calculated, when searches were performed with a Percolator score cut-off of 13 and a significance threshold of  $p < 0.05$ . The spectral files and search results were uploaded to the ProteomeXchange platform (<http://www.proteomexchange.org>) and have the identifier PXD001432. The peptide assignments were re-imported into Progenesis LC-MS and the normalized abundances of all unique peptides were summed up and allocated to the respective protein (Supplemental Table S1 and S2).

#### 2.1.5 Data processing and identification of proteins

The data processing and identification of proteins<sup>157</sup> was performed in cooperation with the Research Unit Protein Science, Helmholtz Center Munich (Dr. Stefanie Hauck and Dr. Christine von Toerne). Proteins were only added to the high confident cell surface protein dataset if one of the following criteria was true: (i) protein was identified in either Trypsin or PNGase fraction with one or more peptides, if the confidence score was  $\geq 18$ , (ii) protein was identified in both fractions (Trypsin and PNGase) with at least one peptide and a confidence score  $\geq 13$ . To guarantee stringency for this analysis, the Ensembl human database protein ID of each identified protein was converted to the respective transcript (ENST) ID. The spectra of single peptide hits were all manually inspected and a complete collection of these spectra can be found in the Supplemental Fig. S1. To furthermore verify the identification of cell surface proteins and to remove potential contaminants, only proteins at least annotated as 'membrane' or 'secreted' in the UniProtKB/Swiss-Prot database<sup>162</sup> were considered for the cell surface protein atlas. The PAL-qLC-MS/MS approach gave the experimental evidence for

the localization of these proteins on the cell surface. The proteins already nomenclatured as CD molecules were also included in the cell surface dataset for further analysis.

#### 2.1.6 Technical validation of PAL-qLC-MS/MS

During the establishment phase of the PAL-qLC-MS/MS technique, the number of cells per sample, the donor-dependency on expression patterns and the influence of the oxidation/biotinylation one-pot-reaction on protein expression patterns was tested. Samples, which were obtained during this establishment phase were prepared and already subjected to LC-MS/MS analysis. The results of this test phase (not shown here) indicated, that it is possible to identify proteins such as CD11a, CD69, CD62L, which are known to be relevant for T cell activation, via the MS analysis. Due to the reason that the oxidation-biotinylation treatment is a harsh chemical process for the cells, a system for a solid technical validation of the LC-MS/MS-based analysis was set up as follows: In parallel to 2.1.1 (Cell surface protein labeling), additional cells of the different donors (D1-D4) at the respective time points during the stimulation time course, were taken to perform a flow cytometry staining (section 2.2) of cell surface proteins which were already identified during the establishment phase of the PAL-qLC-MS/MS technique. This additional analysis created the possibility to be able to compare the protein expression pattern obtained by oxidation-biotinylation, followed by MS analysis and the pattern, which was monitored by a flow cytometry staining. Protein abundances obtained by qLC-MS/MS and mean fluorescence intensity (MFI) monitored by flow cytometry of the single donors at the respective time points were compared. In addition, the overall protein expression pattern of the cells of all donors during the stimulation time course was compared. The via PAL-qLC-MS/MS technique obtained protein abundances from all donors at all time points were subjected to a Principal Component Analysis (PCA) with the add-in software XLStat (add-in software for Microsoft Excel), to investigate the comparability between the different samples.

#### 2.2 Protocols for cell staining and flow cytometry analysis

20,000-150,000 cells per well/per staining were transferred in a 96-well round-bottom plate and centrifuged (805 x g for 1min) to pellet the cells. The supernatant was discarded and the cells were washed once with 200  $\mu$ L flow wash buffer (PBS containing 5 % FCS and 0.02 %  $\text{NaN}_3$ ). The single stain antibodies or antibody staining mixes were prepared on ice and antibodies were diluted in flow wash buffer. The staining was carried out in a total volume of 20  $\mu$ L per well at 4 °C in the dark for 30min. Then, the cells were washed once with flow wash buffer and the cell pellet was resuspended in 200-250  $\mu$ L flow wash buffer. 10  $\mu$ L propidium

iodide was added per staining as live/dead indicator, immediately before recording the samples on a BD LSRFortessa flow cytometer in combination with the software BD FACSDIVA 7.0 and analysis with FlowJo Software. If possible, 10,000-20,000 events were recorded per sample and propidium iodide positive cells were excluded from the analysis. Cells were measured in single-tubes, except for the flow cytometry-based cell surface screening (section 2.2.2) and the testing of hybridoma supernatants (section 3.1.4), for these experiments the HTS Plate Loader unit of the BD LSRFortessa was used for acquisition.

If Aqua staining was used as the live/dead indicator, the cells were washed twice with ice-cold PBS and then stained with 100  $\mu$ L Aqua (1:1000 in ice-cold PBS) at 4 °C in the dark for 30min, before the staining with the respective antibodies. Aqua positive cells were then excluded from the analysis.

If cells needed to be fixed, they were washed once with flow wash buffer after the staining. Then the cells were incubated with 100  $\mu$ L Fix-Solution (BD Cytfix/Cytoperm Kit) at 4 °C in the dark for 20min. The cells were washed twice with 200  $\mu$ L of flow wash buffer and kept at 4 °C in the dark until the measurement.

If a multicolor staining was applied to the sample, a compensation experiment was set up in advance. Therefore, per antibody, a mixture of 100  $\mu$ L flow wash buffer with one drop of BD Comp Beads (Anti-Mouse Ig,  $\kappa$ ; Anti-Rat Ig,  $\kappa$ ; depending on the subtype of the antibody), one drop of Negative control beads and 2-3  $\mu$ L antibody was prepared and incubated at RT in the dark for 30min. 2 mL flow wash buffer was added and bead-antibody mixtures were centrifuged at 200 x g for 10min. The bead pellets were then resuspended in 200  $\mu$ L fresh flow wash buffer each and compensation beads were recorded in the compensation setup. If Aqua was used, one drop of ArC Amine reactive beads was allowed to adjust to RT for 5min in a tube. 2  $\mu$ L of Aqua was added, the mixture was vortexed and incubated at RT in the dark for 30min. One drop of Negative Control Beads and 2 mL of flow wash buffer was added and then centrifuged at 200 x g for 10min. The bead pellet was resuspended in 200  $\mu$ L of fresh flow wash buffer. The final compensation was then calculated with FACSDIVA 7.0 software and could be applied for multicolor experiments.

### 2.2.1 Detailed settings for individual experiments using flow cytometry

Staining of PBMCs was performed on a cell number of 150,000 cells/staining. If isolated T cells were stained for flow cytometry, the number of cells was adjusted to 100,000 cells/staining. Except the flow cytometry-based cell surface screening (see 2.2.2), consisting of only one conjugated antibody per staining and due to the limitation of the total available cell number of one blood donor, the number of cells was adjusted to 20,000 cells/staining.

Purity and viability of naive T cells after the isolation process was always checked via CD45RA<sup>+</sup>/CD45RO<sup>-</sup> staining of the resulting cell population. During the activation of the naive CD4<sup>+</sup> T cells via anti-CD3/anti-CD28 stimulation the activation marker CD69 was monitored to ensure proper activation. The efficiency of oxidation/biotinylation reaction (PAL-qLC-MS/MS) (section 2.1.1) was measured via Streptavidin-PE antibody staining. Detailed information for the respective antibodies is listed in Table M2.

### 2.2.2 Flow cytometry-based cell surface screening and data analysis

This experiment was performed in cooperation with the Institute of Virology, Helmholtz Center Munich (Prof. Dr. Michael Schindler and Dr. Herwig Koppensteiner). A targeted cell surface antigen screening, based on monoclonal antibodies and flow cytometry, was performed with the LEGENDScreen Human Cell Screening (PE) Kit. This kit includes 96 well plates, which are pre-coated with one lyophilized monoclonal PE-conjugated antibody per well. It contains 332 antibodies against human cell surface antigens and 10 Ig isotype controls (mouse, rat, hamster). Naive CD4<sup>+</sup> T cells from 3 different human donors (n=3, D5-D7) were taken in their naive form and in addition cells were activated with anti-CD3/anti-CD28 for 3 and 24h for this experimental setting, generating 9 samples (cells of 3 donors, 3 time points each) for analysis in total. The cells were washed with cold PBS and centrifuged at 290 x g for 10min directly after isolation, respectively after the end of the stimulation. The cells were resuspended in Cell Staining Buffer, included in the kit, and kept on ice. The antibody-coated 96-well plates were centrifuged at 805 x g for 5min, followed by dissolving the antibodies in 75 µL of ddH<sub>2</sub>O. The plates were then incubated at RT in the dark for 15min. After incubation, two times 25 µL were transferred from the original 96-well plate to new 96-well plates. 20,000 cells in 75µL volume were then transferred to each well containing 25 µL of dissolved antibody, resuspended and incubated at 4°C in the dark for 30min. The cells were washed once with 200 µL cell staining buffer per well and then resuspended in 100 µL fixation buffer, provided in the kit. The cells were incubated at RT in the dark for 15min and then washed again with 200 µL cell staining buffer. The fixed cells were resuspended in 200 µL cell staining buffer and kept at 4 °C in the dark until data acquisition.

The resulting data was analyzed using BD FACSDIVA 7.0 and FlowJo Software. Antibody signals were considered as positive when the MFI was higher than the highest measured Ig isotype control. In addition, it was necessary to obtain a positive signal on the cells of at least two donors (otherwise stated in Table 2), if the protein was considered for the cell surface protein atlas.

## 2.3 Analysis of proteomic results: unsupervised clustering by GProx and Gene ontology (GO) enrichment analysis and DrugBank target search

The mean ratio of the protein abundance of the four donors between the time points per protein, which was detected via PAL-qLC-MS/MS, was calculated. These ratios were taken to perform an unsupervised cluster analysis. Regarding the flow cytometry-based cell surface screening, the mean fluorescence intensity ratio between the measured time points of the single proteins of one representative donor, were subjected to unsupervised cluster analysis. The ratios were uploaded as .txt file into the GProx (The Graphical Proteomics Data Explorer) open access software, which is based on the fuzzy c-means algorithm as implemented in the Mfuzz package.<sup>163,164</sup> The graphical output of this software shows graphs of the proteins depending on their expression pattern. It also calculates a membership value for the single proteins per cluster, which indicates how exactly the expression of one protein fits to the general expression pattern of the cluster. For the resulting expression clusters of the PAL-qLC-MS/MS ratios, a Gene Ontology enrichment analysis<sup>165</sup> was performed. The gene names of the proteins of one cluster, which reached at least a membership value of  $\geq 0.6$  regarding this cluster, were uploaded as group to the website of the Generic Gene Ontology (GO) Term Finder.<sup>166</sup> To summarize and visualize these results, the resulting GO terms of one cluster were directly transferred to the web-based software REVIGO.<sup>167</sup> Within this software, the allowed similarity was set to the predefined value “small”, to guarantee that possible GO term pairs will be displayed with a semantic similarity less than 0.5.

All proteins, which were identified via PAL-qLC-MS/MS and/or the flow cytometry screen, were searched as drug targets at the free DrugBank database<sup>168</sup>, using their assigned UniProtKB accession numbers. Due to the high number of Drug target hits within the group of solute carrier proteins (SLCs), this group was especially highlighted in Table 4, Fig. 23 and Fig. 24, using the “R programming language” ([www.r-project.org](http://www.r-project.org)) and the “gplots” package for data analysis (in cooperation with the Institute of Computational Biology, Helmholtz Center Munich, Linda Krause).

## 2.4 Transcriptomic analysis of naive and activated CD4<sup>+</sup> T cells

### 2.4.1 RNA isolation and RNA quality measurement

Naive CD4<sup>+</sup> T cells of four human blood donors (D8-D11) were isolated (see 1.2).  $1 \times 10^6$  naive T cells per donor were activated with anti-CD3/anti-CD28 for 3h (see 1.2) and  $1 \times 10^6$  naive T cells per donor were directly resuspended in 350  $\mu$ L RLT buffer after cell isolation and frozen at -80 °C until RNA isolation. The T cells which were activated after cell isolation, were

resuspended in the 24-well plate, transferred to a reaction tube, centrifuged for 10min at 6000 x g, and the cell pellet was resuspended in 350  $\mu$ L RLT buffer and also stored at -80 °C until RNA isolation. The RNA isolation was performed according to the manufacturer's protocol using the QiaShredder columns and the RNeasy Mini Kit. RNA amount was measured using the NanoDrop spectrophotometer and RNA quality was measured with the RNA 6000 Nano Kit combined with the Agilent 2100 Bioanalyzer, according to the manufacturer's protocol. The obtained RNA Integrity Numbers (RIN) of the samples were  $\geq 9$ .

#### 2.4.2 Whole genome microarray and analysis of resulting transcriptomic data

25  $\mu$ g of total RNA from naive and 3h activated CD4<sup>+</sup> T cells of four human blood donors (D8-D11) was amplified and Cy3-labeled according to the manufacturer's protocol using the 1-color Low Input Quick Amp Labeling Kit. The Hybridization Kit was used to hybridize the samples to the SurePrint G3 Human Gene Expression 8x60K microarray. The transcriptomic dataset was deposited in the NCBI Gene Expression Omnibus (GEO)<sup>169</sup> and is accessible through GEO Series accession number GSE61983. Raw data was imported in the GeneSpring software GX 12.5 and quality control was performed according to the software guidelines. Transcripts, which were detected in the transcriptomic analysis of the samples of all four donors either in the naive T cells, the 3h stimulated T cells or under both conditions were extracted from the GeneSpring software for further analysis.

#### 2.4.3 Sequence mapping, *in silico* identification of transcripts coding for cell surface proteins and GO term analysis

This analysis was performed in cooperation with the Department Informatics, Bioinformatics & Computational Biology i12, Technical University Munich (Dr. Edda Kloppmann and Tatyana Goldberg). To identify transcripts coding for cell surface proteins or proteins which are close to the plasma membrane, an *in silico* analysis was performed. Of a total number of 27,958 human Entrez Gene RNAs, 17,757 microarray probe names were assigned to 14,455 unique NCBI RefSeq<sup>170</sup> accession numbers by Agilent Technologies. 13,028 of these RefSeq accession numbers were mapped to their corresponding UniProtKB accession numbers (AC; UniProt release 2013\_10). It is of note that several human UniProt ACs can be assigned to the same gene name, but to reduce redundancy regarding this fact, the following rule was set: If available, the reviewed UniProt AC (Swiss-Prot) was chosen, if not the unreviewed AC (TrEMBL) was taken into account. Therefore, UniProt ACs for 12,263 gene names (Supplemental Table S3) could be assigned. After this processing, a remaining redundancy of 29 gene names was still present, relying on the fact that these 29 gene names were assigned to more than one reviewed UniProt AC. In addition to these 29, we kept 36 unreviewed UniProt

ACs, based on their prediction as plasma membrane proteins by the following criteria: If available, the subcellular localization (UniProt\_SL) annotation was extracted from the UniProtKB/Swiss-Prot database.<sup>162</sup> If there was no annotation available, the subcellular localization was predicted by the software LocTree3<sup>171</sup> in combination with the software PolyPhobius<sup>172</sup>, which predicts transmembrane helices (TMHs). Proteins were identified as cell surface proteins as follows (Supplemental Table S3): (a) all experimentally verified and probable subcellular localizations from UniProt\_SL were accepted. Proteins annotated as cell surface and cell membrane (plasma membrane) were kept; proteins annotated as localized in the plasma membrane were only retained if they were single- or multi-pass membrane proteins. If no further information for the proteins localized in the plasma membrane was available, an additional criterion was that PolyPhobius needed to predict at least one TMH. If no TMHs were predicted, these proteins were grouped in the section “putative cell surface proteins”. Membrane proteins, which were annotated as peripheral or lipid-anchored, needed to fulfill an additional annotation as localized on the extracellular side of the plasma membrane. Proteins annotated as located on the cytoplasmic side were directly excluded, if no further information on the localized side was available, the proteins were also assigned to the group of putative cell surface proteins. (b) If the localization of the protein could not be annotated due to UniProt\_SL annotation the subcellular localization was predicted by LocTree3, and in addition at least one TMH needed to be predicted by PolyPhobius to classify this protein as plasma membrane protein. A list containing all genes passing these criteria were reimported into GeneSpring GX 12.5 software to analyze their transcripts regarding differential gene expression between the samples of naive T cells and 3h activated T cells. Therefore, a paired Student t-test filtered for a corrected p-value ( $p \leq 0.05$ , Benjamini-Hochberg correction) was applied.

A Gene Ontology enrichment analysis was performed on the dataset containing the cell surface protein coding transcripts. All gene names were uploaded as a group to the website of the Generic Gene Ontology (GO) Term Finder. To summarize and visualize the obtained results, the GO terms were directly transferred to the software REVIGO. Within this software the allowed similarity was set to the predefined value “small”, to guarantee that possible GO term pairs will be displayed with a semantic similarity less than 0.5.

## 2.5 Combination and correlation of proteomic and transcriptomic datasets

The combination and correlation of proteomic and transcriptomic datasets was performed in cooperation with the Institute of Computational Biology, Helmholtz Center Munich (Dr. Bettina Knapp and Linda Krause). The “R programming language”<sup>173</sup> and the “gplots” package were used for data analysis. To be able to combine the omics datasets of the transcriptomic

and proteomic analysis, the datasets needed to be scaled. Therefore, the mean protein abundance of the PAL-qLC-MS/MS experiment, respectively the mean of the measured MFIs of the flow cytometry screen and the corresponding standard deviations were calculated. The mean was then subtracted from each single value of the corresponding measurement and then divided by the calculated standard deviation. The transcriptomic gene expression values, which resulted from the microarray measurement, were log<sub>2</sub> transformed as normalization. These scaled datasets were comparable and a two-sided Welch t-test ( $p \leq 0.01$ , scaled absolute expression measurement of  $\geq 1$ ) was applied. Differentially expressed genes/proteins are presented as a heat map.

The correlation of the transcriptomic dataset to the proteomic dataset obtained via PAL-qLC-MS/MS was calculated on the normalized values for gene expression and protein abundance, using the "R programming language".<sup>173</sup> The correlation analysis was only performed on the targets, which could be measured in all four donors via PAL-qLC-MS/MS as well as RNA level, resulting in 159 correlated targets. The dataset does not come from a normal distribution, so the rank-based Spearman's rho statistic was used as a robust way to estimate the magnitude of association. Corresponding p-values describe the probability of observing a correlation of this extent (or extremer) under the assumption that there is no correlation. P-values were adjusted using the procedure by Benjamini and Hochberg, controlling the false discovery rate.

## **2.6 Identification and targeted validation of cell surface proteins which were not mentioned in the context of T cell biology before**

To extract interesting candidates from the cell surface atlas for further studies, an extensive literature and patent search on <http://www.ncbi.nlm.nih.gov/pubmed> as well as on <http://www.google.com/patents> was performed. The following keywords were used alone and in combination: T cell [OR] immune cell [AND] activation, differentiation, proliferation [AND] the recommended protein name listed at UniProtKB as well as at least two of the alternative names for the protein listed at UniProtKB.

### **2.6.1 Validation of candidates by qPCR**

Naive CD4<sup>+</sup> T cells were isolated from four further human blood donors (D12-D15) as described in section 1.2 and total RNA was isolated (section 2.4.1). The RNA concentration was measured using the NanoDrop 2000 and cDNA was synthesized in a volume of 20  $\mu$ L by using the High-capacity cDNA Reverse Transcription Kit, according to manufacturer's protocol. The exon-spanning qPCR primers with an estimated  $T_m$  at around 58 °C were taken from the

qPrimerDepot (<https://primerdepot.nci.nih.gov>) or designed by using Primer3web (<http://primer3.ut.ee>; <sup>174</sup>) and ordered at Metabion International AG, all primers are listed in Material Table M 5. Each qPCR reaction was composed of 30 ng of cDNA, 5  $\mu$ L of FastStart Universal SYBR Green Mastermix 2 x (ROX), and each forward and reverse primer at a final concentration of 400 nM, filled up to a total volume of 10  $\mu$ L per reaction with DEPC-treated H<sub>2</sub>O. The reactions were carried out in 384-well plates on a Viiia7 Real-Time PCR System according to the following protocol: 50 °C (2min), 95 °C (10min); [95 °C (15sec), 60 °C (1min)] x 40 cycles; melt curve: 95 °C (15sec), 60 °C (1 min), (ramp: 0.05 °C/s), 95 °C (15sec). The amplifications were carried out in at least technical duplicates and EF-1alpha was used as control. If relative changes in gene expression were analyzed, the comparative C<sub>T</sub> ( $2^{-\Delta\Delta C_T}$ ) method was used.

### 2.6.2 Validation of candidates by Western blot

Naive CD4<sup>+</sup> T cells of four additional blood donors (D16-D19) were collected (section 1.2) and protein lysates were prepared for Western blot analysis. Equal protein amounts in lysis buffer (1 % NP-40, 10 mM NaCl, 10 mM Tris pH 7.6, 2 x EDTA-free complete protease inhibitor cocktail in ddH<sub>2</sub>O)<sup>157</sup> were incubated with NuPAGE LDS Sample Buffer and heated at 95 °C for 10min, loaded on 10 % Bis-Tris protein gels or 4-12 % Bis-Tris Protein Gels and separated for 1.5h at 120 V using SDS-PAGE gelelectrophoresis in a MOPS SDS running buffer. The proteins were transferred to a polyvinylidenfluorid (PVDF) membrane using XCell I Blot Module at 60 V (limited to 500 mA) for 90min. To block unspecific binding sites, an incubation step in 3-5 % non-fat dry milk in 1 x PBS for one hour was carried out. The primary antibodies (Material Table M 3) were diluted in 3-5 % non-fat dry milk in PBS in a 50 mL reaction tube and the membranes were incubated at 4 °C overnight, rolling. Three washing steps with 3-5 % non-fat dry milk in PBS (10min each) were done before incubation with the corresponding secondary HRP-linked antibody (Material Table M 4) at 4 °C for 2h, rolling. The membranes were washed with PBS containing 0.02 % Tween for three times and then incubated with the HRP-substrate Amersham ECL Prime Western Blotting Detection Reagent for 5min. The chemiluminescence of the HRP reaction on all membranes was recorded with an ECL ChemoCam Imager and ChemoStar software. A positive Western blot assay showed a band at the expected molecular weight of a protein.

### **3. Analysis of transmembrane protein c16orf54, a cell surface protein which was not described in the context of T cell biology before**

The transmembrane protein c16orf54 (named c16orf54) is an uncharacterized human protein, which is expressed on human naive CD4<sup>+</sup> T cells and is encoded by the chromosomal open reading frame 54 on chromosome 16. To investigate this target protein in detail, tools and expression systems needed to be generated.

#### **3.1 Generation, production and testing of rat and mouse monoclonal antibodies against human and murine c16orf54**

The generation of a monoclonal antibody against c16orf54 was carried out in cooperation with the Institute of Molecular Immunology (IMI) of the Helmholtz Center Munich (Dr. Elisabeth Kremmer and Dr. Regina Feederle). Peptides for the immunization of mice and rats were chosen based on the extracellular c16orf54 amino acid sequence of human (ORF54S) and murine (ORF54M) origin, according to the UniProtKB. For immunization, peptides were coupled via a terminal cysteine to maleimide-activated-ovalbumin. An internal cysteine residue was exchanged against alpha butyric acid (Abu) to avoid cross-coupling of ovalbumin. Biotinylated peptides served as antigens for antibody testing. The peptides were purchased from Peps4LS (Heidelberg, Germany):

- 1) ORF54S-Cys (MPLTPEPPSGRVEGPPAWEAAPWPSLP-Abu-GP-C)
- 2) ORF54S-eBio (MPLTPEPPSGRVEGPPAWEAAPWPSLP-Abu-GP-Spacer-Biotin)
- 3) ORF54M (MPVTPQQPSGHTTEGLPEPTAEAAVWVVIP-C)
- 4) ORF54M-eBio (MPVTPQQPSGHTTEGLPEPTAEAAVWVVIP-Spacer-Biotin)

##### **3.1.1 Peptide-immunization of rats and mice and hybridoma generation**

The IMI applied their standard protocol for the hybridoma cell generation.<sup>175,176</sup> In brief, to immunize animals for B cell clone production, approximately 40 µg of the ovalbumin-coupled peptides were emulsified in an equal volume of incomplete Freund's adjuvant combined with 5 nmol CpG2006, and injected subcutaneously and intraperitoneally into two different Lou/C rats and one C57BL/6J mouse. The immune response was boosted intraperitoneally and subcutaneously after six weeks with 40 µg peptides without Freund's adjuvant. Three days after the boosting injection, cells of the myeloma cell line P3X63-Ag8.653 were fused with rat or mouse spleen cells according to the standard procedure described by Köhler and Milstein.<sup>177</sup> The fused hybridoma cells were cultured in 96-well plates with RPMI

1640 medium (supplemented with 20 % FCS, 100 U/mL Pen/Strep, 1 % sodium pyruvate, 1 % non-essential amino acids, 2 % hybridoma cloning factor, 2 % hypoxanthin-aminopterin supplement).

### 3.1.2 Testing of hybridoma supernatants against biotinylated peptides via ELISA

Hybridoma supernatants were tested by the IMI of the Helmholtz Center Munich in a solid-phase enzyme-linked immunoassay (ELISA) two weeks after fusion. Biotinylated peptides were bound to avidin-coated 96-well plates at a concentration of 0.2 µg/mL in 0.1 M sodium carbonate buffer (pH 9.6) at 4 °C overnight. Irrelevant biotinylated peptides were used as negative controls. After blocking with PBS (containing 2 % FCS) for one hour at RT, the hybridoma supernatants were added and incubated for 30min. Rat monoclonal antibodies from the hybridoma supernatants, which bound to the coated peptides, were detected by using a mixture of HRP-conjugated monoclonal mouse-anti-rat IgG heavy chain antibodies (anti-IgG1, anti-IgG2a, anti-IgG2b, anti-IgG2c). Respectively, mouse monoclonal antibodies from the hybridoma supernatants, which bound to the coated peptides, were detected by using a mixture of HRP-conjugated monoclonal rat-anti-mouse IgG heavy chain antibodies (anti-IgG1, anti-IgG2a, anti-IgG2b, anti-IgG3). HRP was visualized with ready to use TMB substrate. Hybridoma clones, which were positive in the ELISA screening for any of the peptides used for immunization (human and/or mouse sequence), were further characterized regarding their IgG isotype in a second ELISA using anti-light chain antibodies as capture and HRP-coupled anti-IgG subclass-specific antibodies for detection. The IgG specified hybridoma supernatants were transferred to the Center of Allergy and Environment for further testing. Hybridoma cells of positively tested supernatants (anti-human-c16orf54 clones 23H8 and 7F11 and anti-murine-c16orf54 clone 13D2; all rat IgG2c) were cloned at least twice by limiting dilution to obtain stable monoclonal hybridoma cell lines. The supernatants of the sub-cloned hybridoma cells were tested again via flow cytometry and the purification of these antibodies was performed on Protein-A-Sepharose 4 Fast Flow columns. Bound IgGs were eluted with 0.1 M citrate buffer, pH 4.0 and dialyzed three times against PBS. The protein concentrations of the purified antibodies were spectrophotometrically determined.

### 3.1.3 Isolation of murine naive CD4<sup>+</sup> T cells

The spleen of sacrificed C57BL/6 wt mice was isolated and the spleens were manually cut into small pieces with a scissor. These small pieces were additionally filtered through a 100 µm cell strainer with the plunger of a syringe to destroy the tissue connection. The cell strainer was flushed with 10 mL D-PBS and cells were pelleted by centrifugation (290 x g, 10min, 4 °C). The pellet was resuspended in 10 mL 1 x Ery-Lysisbuffer (0.15 M NH<sub>4</sub>Cl; 10 mM

KHCO<sub>3</sub>; 0.1 mM Na<sub>2</sub>EDTA; pH 7.2-7.4) per spleen and incubated for 6min at RT. The lysis reaction was stopped by addition of 10 mL RPMI 1640 medium containing 10 % FCS. The cells were centrifuged (290 x g, 10min, 4 °C), resuspended in 10 mL autoMACS running buffer per spleen and filtered through a new 100 µm cell strainer. The resulting cell population of splenocytes was counted (Neubauer counting chamber, see 1.2) and pelleted via centrifugation (290 x g, 10min, 4 °C). To isolate murine naive CD4<sup>+</sup> T cells out of the splenocytes the CD4<sup>+</sup> CD62L<sup>+</sup> T cell Isolation Kit II (mouse) was used. The splenocytes were resuspended in autoMACS running buffer (400 µL autoMACS running buffer per 10<sup>8</sup> cells) and incubated with the Biotin Antibody Cocktail II (100 µL Biotin Antibody Cocktail II per 10<sup>8</sup> cells) for 10min on ice. AutoMACS running buffer (300 µL per 10<sup>8</sup> cells) and anti-biotin microbeads (200 µL per 10<sup>8</sup> cells) were added and incubated for 15min on ice. 10 mL of autoMACS running buffer were added and the cells were centrifuged (290 x g, 10min, 4 °C). The cell pellet was resuspended in autoMACS running buffer (500 µL autoMACS running buffer per 10<sup>8</sup> cells) and CD4<sup>+</sup> T cells were isolated by negative selection through an automated magnetic column (autoMACS, program: DEPLETES). To specifically enrich the naive CD4<sup>+</sup>/CD62L<sup>+</sup> T cells from this negative selected population, the cells were counted, pelleted and resuspended in autoMACS running buffer (800 µL autoMACS running buffer per 10<sup>8</sup> cells). CD62L microbeads were added (200 µL per 10<sup>8</sup> cells) and incubated for 15min on ice. After the addition of 10 mL autoMACS running buffer and centrifugation (290 x g, 10min, 4°C), the cells were resuspended in autoMACS running buffer (500 µL autoMACS running buffer per 10<sup>8</sup> cells). Naive CD4<sup>+</sup>/CD62L<sup>+</sup> T cells were enriched by positive selection (autoMACS, program: POSSEL) of the labeled cells.

#### 3.1.4 Suitability test of hybridoma supernatants (flow cytometry and Western blot)

The via ELISA positive tested hybridoma supernatants were evaluated based on their suitability for the application in flow cytometry and Western blot at the Center of Allergy and Environment. Information for each hybridoma supernatant included the specific antibody isotype contained in the supernatant as well as the information if the hybridoma supernatant reacted against the human-sequence-, the murine-sequence- or both peptides. Therefore, murine and human naive CD4<sup>+</sup> T cells were isolated (section 1.2 and 3.1.3) and stained for flow cytometry (see section 2.2) with the respective species reacting hybridoma supernatants. For the case of the hybridoma supernatants originating from the fusion with rat spleen cells, the following 3-step staining was performed, all steps on ice and each incubation step for 30min. 100,000 cells per well were seeded in a 96-well plate and washed with flow wash buffer. The cells were first stained with 50 µL of the anti-c16orf54 hybridoma supernatants and then washed once. The second staining step was a 1:10 dilution of hybridoma superna-

tant containing mouse-anti-rat specific antibodies for the indicated isotype of the antibody of the anti-c16orf54 hybridoma supernatant (anti-rat-IgG1, anti-rat-IgG2a, anti-rat-IgG2c). After a washing step, the cells were stained with rat-anti-mouse IgG (H+L) AlexaFlour594 (1:200). In the case of the hybridoma supernatants originating from the fusion with mouse spleen cells, the following 3-step staining was performed, all steps on ice and each incubation step for 30min. 100,000 cells per well were seeded in a 96-well plate and washed with flow wash buffer. The cells were first stained with 50  $\mu$ L of the anti-c16orf54 hybridoma supernatants and then washed once. The second staining step was a 1:10 dilution of hybridoma supernatant containing rat-anti-mouse specific antibodies for the indicated isotype of the antibody of the anti-c16orf54 hybridoma supernatant (anti-mouse-IgG1, anti-mouse-IgG2a, anti-mouse-IgG2b). After another washing step, the cells were stained with mouse-anti-rat DyLight 594 AffiniPure IgG (H+L) (1:200). As a negative control the cells were stained with a hybridoma supernatant containing antibodies against a protein, which is not expressed on T cells.

Selected hybridoma supernatants, which were suitable for the application in flow cytometry on human naive CD4<sup>+</sup> T cells, were also examined about their suitability in Western blot technique. The general settings for this technique are described in section 2.6.2. Hybridoma supernatants were tested on cell lysates of human naive CD4<sup>+</sup> T cells, respectively HEK-293 cell lysates. After blotting the proteins onto the PVDF membrane the single lanes were separated from each other to probe the stripes with different hybridoma supernatants (1:10 dilution). For the second incubation of the stripes, isotype specific HRP-conjugated antibodies were chosen (1:1000).

### 3.1.5 Direct labeling of antibody and antibody-peptide competition assay

The purified monoclonal antibody anti-human c16orf54 of clone 23H8\* was chosen for direct labeling with a fluorophore for the use in a flow cytometry staining. Therefore, the Lightning-Link PE-TexasRed Tandem Conjugation Kit was used. 60  $\mu$ g anti-human c16orf54 23H8\* were diluted with PBS to a total volume of 40  $\mu$ L and 5  $\mu$ L LL-modifier were added. This mixture was transferred to the glass vial containing the lyophilized fluorophore PE-TexasRed, carefully mixed and incubated at 4 °C in the dark overnight. The next day, 5  $\mu$ L LL-quencher were added and the mix was incubated at 4 °C in the dark for further 30min. The resulting concentration of this labeled antibody, called anti-human c16orf54 23H8\* PE-TexasRed (abbreviated as 23H8\*-PE-TexasRed) in the following sections, was 1 mg/mL. To assess the working dilution of the labeled antibody, naive T cells and PBMCs were stained with different concentrations together with Aqua as live/dead indicator (section 2.2).

To test the specificity of the labeled antibody for flow cytometry applications, a competition assay was carried out. Therefore, equal amounts of c16orf54 23H8\* PE-TexasRed and the peptide, which was used for immunization of the rats, as well as excess of the peptide (antibody:peptide → 1:2-1:400) were diluted in flow wash buffer in a volume of 10 µL and incubated at RT in the dark for 30min. Freshly isolated PBMCs were then incubated in this mixture at 4 °C in the dark for 30min and measured via flow cytometry.

### **3.2 Generation of expression systems for recombinant expression of c16orf54**

To create cellular model systems constantly expressing c16orf54, different expression vectors were cloned and the cells were stably transfected with the different constructs.

#### **3.2.1 Cloning of different overexpression vectors**

Molecular Biology standard procedures such as DNA restriction digestion, ligation, PCR, plasmid isolation and transformation were performed according to established protocols.<sup>178</sup>

##### **3.2.1.1 Amplification of inserts**

RNA from  $38 \times 10^6$  naive CD4<sup>+</sup> T cells of one human blood donor was isolated and transcribed into cDNA (section 2.4.1 and 2.6.1). To amplify the coding region, 80 ng cDNA were used for each of the five PCR reactions with the following primer combinations (sequences in Material Table M 5): c16orf54\_P1 + c16orf54\_P2 resulting in construct c16orf54\_1; c16orf54\_P3 + c16orf54\_P2 resulting in construct c16orf54\_2; c16orf54\_P1 + c16orf54\_P4 resulting in construct c16orf54\_3; c16orf54\_P5 + c16orf54\_P6 resulting in construct c16orf54\_4; c16orf54\_P7 + c16orf54\_P6 resulting in construct c16orf54\_5 (10 µM of each primer per reaction) (overview of constructs Fig. 30 in section: Results). The designed primers included restriction sites for either NheI or XhoI (in case of constructs c16orf54\_1, \_2 and \_3) and restriction sites for XbaI or NotI (in case of constructs c16orf54\_4 and \_5).

The Accu Prime Pfx DNA Polymerase was used according to manufacturer's recommendations for the amplification in a volume of 25 µL. The PCR protocol was programmed with the following conditions: 95 °C (2min); [95 °C (15sec), 55 °C (30sec), 68 °C (45sec)] x 34 cycles; 68 °C (5min), 4 °C hold. The PCR products were loaded with 6 x Loading Dye and ran on a 1.5 % agarose gel. Bands with correct size were excised from the gel and cleaned up by using the GeneJet Gel Extraction Kit according to manufacturer's recommendation. Each of the amplicon samples was eluted in 40 µL of ddH<sub>2</sub>O.

### 3.2.1.2 Ligation and Transformation

To digest the whole amount of the amplicons with the corresponding restriction enzymes to produce the needed overhangs, c16orf54\_1, \_2, \_3 and 1 µg of pcDNA3.1 were each incubated with 1 µL NheI (Fast Digest) and 1 µL XhoI (Fast Digest) and 5 µL 10 x Fast Digest Green Buffer. c16orf54\_4, \_5 and 1 µg pAcGP67-B were each incubated with 1 µL XbaI (Fast Digest) and 1 µL NotI (Fast Digest) and 5 µL 10 x Fast Digest Green Buffer. The restriction reaction mixes were incubated at 37 °C for 30min. The mixes were loaded on a 1.5 % agarose gel. The bands were excised and the DNA was purified with the GeneJet Gel Extraction Kit according to manufacturer's recommendation and samples were eluted in 40 µL ddH<sub>2</sub>O. To ligate the purified and digested inserts of interest into their corresponding vectors, the DNA amounts were estimated according to the thickness of the band in the agarose gel before and for the ligation reaction a ratio of vector to insert of 4:1 was used. 1 µL (5U) T4 DNA Ligase, 2 µL 10 x Ligase Buffer, ddH<sub>2</sub>O, vector and insert were mixed and incubated at RT for 2h.

To transform bacteria with the ligated vectors, 50 µL of chemo-competent NEB5alpha, per ligation reaction mix, were thawed on ice for 10min. 10 µL of each ligation reaction mix was added to the thawed bacteria and this was mixed by pipetting up and down and incubated for further 30min on ice. Heat shock was performed at 42 °C for 30sec and the transformation mix was incubated for 5min on ice. 950 µL of SOC-medium were added and samples were incubated at 37 °C, shaking at 225 rpm for 30min. 50 µL of each transformation were then spread on one Ampicillin (100 µg/mL) containing LB-agarose-plate each, which were then incubated at 37 °C overnight.

### 3.2.1.3 Colony PCR and isolation of plasmids

To check whether the plasmid inside the different bacteria clones is correct, a colony-PCR was performed. Therefore, 5-10 colonies per plate were picked with pipette tips by hand and each resuspended in 20 µL of ddH<sub>2</sub>O. 0.5 µL of this cell suspension were mixed with the T7\_Promoter\_fw (c16orf54\_1, 2, 3) or pAcGP67-B\_fw (c16orf54\_4, \_5), the construct corresponding gene-specific reverse primer (10 µM of each primer per reaction), 10 mM dNTP-Mix, 10 x Dream Taq buffer and 0.2 µL DreamTaq DNA Polymerase in a total volume of 30 µL. The PCR protocol was programmed with the following conditions: 95 °C (3min); [95 °C (30sec), 55 °C (30sec), 72 °C (1min)] x 34 cycles; 72 °C (10min), 4 °C hold. The PCR products were loaded with 6 x Loading Dye and ran on a 1.5 % agarose gel. The clones of each construct, which showed a band on the gel with the correct size, were chosen for amplification. Therefore, the cell suspensions (picked for colony PCR) of the correct clones were inoculated in 5 mL of LB-medium (containing 100 µg/mL Ampicillin) and cultured at 37 °C, shaking at

225 rpm, overnight. To prep the plasmids, the GeneJet Plasmid Miniprep Kit was used, according to manufacturer's recommendation. The plasmid concentration was measured with the NanoDrop spectrophotometer and the correctness of the sequence was confirmed by sequencing at the company GATC Biotech (Konstanz, Germany).

### 3.2.2 Expression of c16orf54 in mammalian cells

To transfect HEK-293 cells with the vectors for the different c16orf54 constructs (c16orf54\_1, 2, 3), the cells were seeded into 6-well plates at a concentration of  $1 \times 10^6$  cells per well and culture medium (DMEM, 10 % FCS, 100 U/mL Pen/Strep) was added up to 4 mL. The cells were cultured at 37 °C and 5 % CO<sub>2</sub> in a humidified incubator overnight. On the next day, the transfection mix for each construct was prepared. 125 µL Opti-MEM were mixed with 7.5 µL Lipofectamine 3000 Reagent for each construct. 1.2 µg of each plasmid DNA from the different constructs was diluted in 125 µL Opti-MEM and 10 µL of P3000 Reagent was added. The 125 µL Lipofectamine-OptiMEM-Mix and the 125 µL DNA-OptiMEM-P3000 Reagent-Mix were combined, cautiously pipetted up and down and incubated at RT for 5min. This transfection mix was added to the seeded HEK-293 cells (6-well plate). Three days after transfection, the HEK-293 cells were trypsinized and all cells were seeded into a T25 culture flask and G418 (400 µg/mL) was added to specifically select stable clones. After three days, the cells were again trypsinized and seeded into a T75 culture flask. Every third day, the cells were split to keep them subconfluent in culture and every third splitting time, G418 (400 µg/mL) was added to preserve the selection pressure for the stable clones. To preserve the stable clones permanently,  $1 \times 10^6$  cells of each construct were resuspended in freezing medium (70 % DMEM, 20 % FCS, 10 % DMSO) and immediately put into a pre-chilled Mr. Frosty, kept at -80 °C for one day and afterwards transferred to the liquid nitrogen tank.

### 3.2.3 Expression and purification of soluble forms of c16orf54 in insect cells

The constructs c16orf54\_4 and \_5 were cloned to be expressed as soluble forms in Sf9 (*Spodoptera frugiperda*) insect cells. Therefore,  $1 \times 10^6$  Sf9 cells were seeded into a T25 culture flask with 5 mL of insect cell medium (Insect Xpress) containing gentamicin (10 µg/mL) at 27 °C. After seeding the cells, they were incubated at 27 °C to become adherent until the transfection mixes were ready to be transfected. The transfection mix 1 contained 2 µg plasmid c16orf54\_4 or \_5 (pAcGP67-B vector), 2.5 µL Baculovirus DNA (0.1 µg/mL, Pro Green) and 100 µL insect cell medium. The transfection mix 2 contained 100 µL insect cell medium and 10 µL Cellfectin II Reagent. Both transfection mixes were combined (one mix for each construct), vigorously mixed for 15s and incubated at RT for 30min. Insect cell medium was added to each mix up to 1 mL total volume, the insect cell medium on the cells was re-

moved and the transfection mix was added. The cells were then incubated for 4h at 27 °C and the flask was slightly shaken every 30 min. 4 mL of insect medium was added and to produce high titer virus stocks, the cells were cultured for four to five days. The supernatant was then centrifuged at 4000 x g for 10min. The clear supernatant, containing the Baculovirus, was used to infect  $2.4 \times 10^7$  Sf9 cells in a T175 culture flask in a total volume of 20 mL of insect cell medium and incubated for five days (first amplification). The whole supernatant was again centrifuged at 4000 x g for 10min. 1 mL of the clear supernatant was again taken to infect  $2.4 \times 10^7$  Sf9 cells in a T175 culture flask in a total volume of 20 mL of insect cell medium and cultured for five days (second amplification). This supernatant was again centrifuged at 4000 x g for 10min and stored at 4 °C in the dark. For protein expression, 1 mL of the cleared supernatant of the second amplification was used to infect  $1.5 \times 10^6$ /mL Sf9 cells in a volume of 400 mL of insect cell medium in a baffled Erlenmeyer flask (volume 2000 mL). To achieve a high protein expression, these cells were then cultured at 27 °C, shaking at 110 rpm for three days. The supernatant was harvested by centrifugation at 4000 x g for 10min and filtered through a 0.45 µm filter. The centrifuged and filtered supernatant was then applied to a nickel-chelating affinity matrix (HisTrap excel, column volume: 1 mL) after washing and equilibrating the purification system (ÄKTA, flow rate 1 mL/min). The column was washed with 5 CV (column volumes) ddH<sub>2</sub>O and then equilibrated with 5 CV PBS (pH 8). The filtered supernatant was applied to the column. The column was pre-washed with PBS pH 8 until the UV chromatogram reached baseline levels. Subsequently, the column was washed with following steps: 5 CV PBS pH 8, with 3 % PBS, 300 mM imidazole (3 % wash), 10 CV PBS pH 8 and a gradient of 3-15 % PBS, 300 mM imidazole (3-15 % wash), 10 CV PBS pH 8 with 15 % PBS, 300 mM imidazole (15 % wash). The recombinantly produced proteins were eluted with PBS pH 8 (containing 300 mM imidazole) in 40 x 1 mL fractions. The different purification fractions were analyzed via SDS-PAGE combined with Coomassie staining<sup>179</sup> (protocol for the self-made SDS gels and Colloidal Coomassie staining can be found in the Material Table M 6). 20 µL of selected fractions (+ 5 µL loading dye) were heated to 95 °C for 10min and 12 µL per fraction were simultaneously loaded on two SDS gels, which were run at 140 V (constant) for 1h 10min. One of the gels was washed with ddH<sub>2</sub>O and then stained with Colloidal Coomassie solution (two days, de-staining with ddH<sub>2</sub>O for 15min) and one gel was semi-dry blotted onto a PVDF membrane (100 mA constant; 1h 30 min) and proteins were detected via anti-V5 antibody or Strep-Tactin-AP conjugate.

### 3.2.4 Immunoprecipitation of recombinant c16orf54

HEK-293 cells ( $1.5 \times 10^6$ ) stably expressing the constructs c16orf54\_2 or \_3 were centrifuged and the cell pellet was resuspended in 500 µL lysis buffer (1 % NP-40, 10 mM NaCl, 10 mM

Tris pH 7.6, 2 x EDTA-free complete protease inhibitor cocktail in ddH<sub>2</sub>O)<sup>157</sup> and stored at -20 °C. After thawing, the lysate was centrifuged at 6000 x g for 10min to remove cell debris. The immunoprecipitation was performed by utilizing the iba system, containing MagStrep type 2HC beads and corresponding buffers. 200 µL of each cleared lysate was incubated with 40 µL pre-washed MagStrep type 2HC beads for 1h at 4 °C, rotating. After this, the reaction tube was placed in a suitable magnet station and the supernatant above the beads was kept as flow-through (FT) fraction. The beads were washed three times with 200 µL of buffer W and after washing, the beads were incubated with 50 µL buffer BE for 5min on ice to elute the bound proteins (fraction IP= immunoprecipitation). The fractions were run on a gel, blotted (section 2.6.2) and probed with anti-c16orf54 23H8\*.

### 3.3 Generation of a CRISPR/CAS mediated knockout mouse

#### 3.3.1 Cloning of vectors containing different sgRNAs

To generate a CRISPR/CAS mediated KO mouse<sup>180</sup>, a strategy containing two sgRNAs targeting upstream and downstream sequences of the transmembrane region of AI467606, was planned. To select target sites for sgRNAs with minimal off-target effects, the extracted genomic sequence of the AI467606 coding exon was subjected to CRISPR Design Tool (<http://crispr.mit.edu/>). The sequences of the ordered oligonucleotides, containing overhangs to BbsI digested pbs-U6\_chimaericRNA, can be found in Table M5. Oligonucleotides were dissolved in TE-Buffer (QIAquick PCR purification Kit) at a concentration of 1 µg/µL. To construct four vectors, each containing one of the sequences for the sgRNA m1 and m3 and the T7\_m1 and T7\_m3, 1 µg of each of the corresponding oligonucleotides: mCRISPR1\_chimA + mCRISPR1\_chimB, mCRISPR3\_chimA + mCRISPR3\_chimB, mT7\_CRISPR1-chimA + mT7\_CRISPR1-chimB, mT7\_CRISPR3-chimA + mT7\_CRISPR3-chimB were combined in four reactions (100 µL total volume in TE-Buffer for each reaction) and incubated at 99 °C for 5min. To generate double strand oligonucleotides, the samples were slowly cooled down to RT by switching off the heating block. To digest the designated vector (pbs-U6\_chimaericRNA)<sup>181</sup>, 5 µg of the vector were incubated with 5 µL 10 x NEB 2.1 buffer and 5 µL BbsI (filled up to 50 µL total volume with ddH<sub>2</sub>O) at 37 °C for 1.5h. The whole volume of the digestion reaction was loaded on a 1.5 % agarose gel and run at 120 V for 30min, the open vector was then gel-purified (section 3.2.1.1) and the concentration was measured via NanoDrop. The annealed oligonucleotides (chimA + chimB) were then ligated with the vector. Four ligation reactions were prepared, each consisting of 100 ng of the digested and gel-purified vector, 1.5 µL of the annealed oligonucleotide-pairs, 1.5 µL 10 x T4 Ligase reaction buffer and 1 µL of T4 DNA Ligase, in a total volume of 15 µL. The ligation reaction was incu-

bated at RT for 1h. To reduce a potential background from re-ligated vector, the ligation reaction was again digested with BbsI: 3  $\mu$ L of NEB 2 buffer, 12  $\mu$ L of ddH<sub>2</sub>O and 1  $\mu$ L of BbsI were added to each ligation reaction, incubated at 37 °C for 20min and heat-inactivated at 65 °C for 20min. 1  $\mu$ L of each ligation reaction was transformed into DH5alpha bacterial cells and plated on Ampicillin containing LB-plates (section 3.2.1.2). On the next day, three colonies per plate (construct) were picked and cultured in 5 mL LB medium (100  $\mu$ g/mL Ampicillin) at 37 °C overnight. Miniprep of these overnight cultures was carried out with the Macherey Nagel Nucleo Spin Plasmid Kit, according to manufacturer's recommendation. The sequences of the vectors were confirmed by sequencing (GATC Biotech). The sequences for the vectors containing mCRISPR1, mCRISPR3 and mT7\_CRISPR3 were correct and a plasmid Maxiprep was performed with the Qiagen Plasmid Maxi Kit, according to the manufacturer's protocol. The concentration and quality of the plasmid preparation was checked via NanoDrop measurement.

### 3.3.2 *In vitro* activity test of the different sgRNAs and T7 endonuclease assay

To test the functionality of sgRNAs mCRISPR1 and mCRISPR3, an *in vitro* experiment with the murine cell line Neuro-2A was set up. Neuro-2A cells were seeded in a 48-well plate (Starlab) at a density of  $5 \times 10^4$  cells in 500  $\mu$ L culture medium (DMEM, 10 % FCS, 100 U/mL Pen/Strep) per well and cultured at 37 °C and 5 % CO<sub>2</sub> in a humidified incubator overnight. On the next day, the medium was changed against 300  $\mu$ L fresh culture medium. The plasmids were prepared in a volume of 30  $\mu$ L Opti-MEM containing (per well): 0.12  $\mu$ g pCas9-T2A-GFP, 0.09  $\mu$ g mCRISPR1 and mCRISPR3. As a negative control, only 0.12  $\mu$ g pCas9-T2A-GFP (no sgRNA containing plasmid) was used for transfection. 0.9  $\mu$ L XtremeGene HP DNA Transfection Reagent was added to each prepared plasmid sample, mixed by careful pipetting and incubated at RT for 20min. 30  $\mu$ L of the transfection mix were added to each well of the prepared Neuro-2A cells in the 48-well plate. Two days later, the DNA of the transfected Neuro-2A cells was isolated with the Promega Wizard Genomic DNA purification Kit, according to manufacturer's protocol. To check the functionality of sgRNAs, a PCR of the target region was performed by using the primer pair CRISPR\_checkP2 and the DreamTaq DNA polymerase (section 3.2.1.3) with the following PCR protocol: 95 °C (3min); [95 °C (30sec), 55 °C (30sec), 72 °C (1min)] x 33 cycles; 72 °C (10min), 4 °C hold. The PCR product was purified by using the GeneJet Gel Extraction Kit according to manufacturer's recommendation. The DNA concentration was measured via NanoDrop and subjected to a T7 endonuclease assay, to check if the sgRNAs cut the DNA in the Neuro-2A cells. Therefore, 600 ng of each sample were mixed with 2  $\mu$ L NEB buffer 2 (in a total volume of 20  $\mu$ L, filled up with ddH<sub>2</sub>O) and incubated in a PCR machine with the following settings: 95 °C (5min),

95-85 °C (ramp: 75 %), 85 °C (30sec), 85-25 °C (ramp: 3 %), 25 °C (30sec), 4 °C hold. 1 µL T7 endonuclease was then added to each of these samples and incubated at 37 °C for 3h. To check whether the T7 endonuclease detected a mismatch, the samples were loaded on a 1.5 % agarose gel and run at 120 V for 1h.

### 3.3.3 Generation of T7-PCR templates for *in vitro* transcription of sgRNAs

To generate T7-PCR templates for the preparation of sgRNAs (T7\_sgRNA\_mCRISPR1 and T7\_sgRNA\_mCRISPR3), two different strategies for m1 and m3 needed to be carried out. The cloning of the vector T7\_m3 was successful (section 3.3.1) and therefore 15 ng of this vector were subjected to one PCR reaction containing 5 x Herculase II buffer, dNTPs (10 mM each), 10 µM Primer\_T7\_Tracr\_1 and Primer\_T7\_Tracr\_2, 2 µL Herculase II, in a total volume of 100 µL. The cloning of T7\_m1 was not successful (section 3.3.1) and therefore a different pair of vectors needed to be used. 15 ng of the vector containing sgRNA mCRISPR1 (mCRISPR1\_chimA + mCRISPR1\_chimB) were mixed with 5 x Herculase II buffer, dNTPs (10 mM each), 10 µM long\_Primer\_m1\_fw and primer Tracr RV, 2 µL Herculase II, in a total volume of 100 µL. The PCR protocol for both reactions was programmed with the following conditions: 98 °C (30sec); [98 °C (5sec), 57 °C (15sec), 72 °C (15sec)] x 45 cycles; 72 °C (5min), 4 °C hold. The size of the resulting PCR products was confirmed by loading 5 µL of the samples on a 2.5 % agarose gel (expected size of the band for T7\_sgRNA\_mCRISPR1: 133 bp; for T7\_sgRNA\_mCRISPR3: 275 bp).

1 µL of the enzyme DpnI was added to each sample and incubated at 37 °C for 30min. The enzyme was inactivated at 80 °C for 10min. After this, the PCR products were purified by using the QIAquick PCR purification Kit combined with the columns of the Qiagen MinElute PCR Purification Kit, according to the manufacturer's recommendation. The samples were each eluted in 10 µL of buffer EB and the concentration was measured with an Eppendorf BioPhotometer.

### 3.3.4 *In vitro* transcription and purification of Cas9 and sgRNAs

To transcribe the T7 templates T7\_sgRNA\_mCRISPR1 and T7\_sgRNA\_mCRISPR3 into the sgRNAs mCRISPR1 and mCRISPR3, the Ambion Mega Short Script Kit was used according to the manufacturer's protocol. 3 µg of each template were used in two separate transcription reactions and they were incubated at 37 °C overnight. After the transcription, 1 µL of Turbo DNase (Ambion Mega Short Script Kit) was added to each reaction and this mix was incubated at 37 °C for 15min to digest the remaining DNA templates. For the purification of the *in vitro* transcribed sgRNAs, the Ambion MEGAclean Kit was utilized according to the manufacturer's recommendation. The purified sgRNAs were resuspended in 40 µL T<sub>10</sub>E<sub>0.1</sub> injection

buffer and incubated at 37 °C for 10min to completely solve them. The concentration was measured on an Eppendorf BioPhotometer and the sgRNAs were stored at -80 °C in aliquots. The Cas9 mRNA preparation was performed by Dr. Oskar Ortiz from the Institute of Developmental Genetics (IDG) of the Helmholtz Center Munich. To prepare Cas9-162A mRNA for the microinjection, 10 µg of the plasmid pCAG-Cas9-162pA were digested at 37 °C for 2h with XbaI (2 µL), AsiSI (4 µL) and AscI (4 µL) in 10 x NEB CutSmart buffer in a total volume of 100 µL. The digestion reaction was run on a 0.9 % agarose gel, the 4.5 kb band was excised and this T7-Cas9-162pA *in vitro* transcription template was purified with the Qiagen gel extraction Kit, according to manufacturer's protocol. The DNA concentration was measured with the Eppendorf BioPhotometer. 1 µg of the T7-Cas9-162pA template was then *in vitro* transcribed by using the Ambion mMESAGE mMACHINE\_T7 Ultra Transcription Kit and purified with the Ambion MEGAClear Kit, according to manufacturer's protocol. The transcribed and purified Cas9 mRNA was dissolved in 30 µL of T<sub>10</sub>E<sub>0.1</sub> injection buffer and the concentration was determined with the Eppendorf BioPhotometer. The quality of sgRNA mCRISPR1, sgRNA mCRISPR3 and Cas9 mRNA was examined by using the Agilent RNA 6000 Nano Kit and samples were analyzed with the Agilent 2100 Bioanalyzer (section 2.4.1).

### 3.3.5 Embryo microinjection, culture and transfer

The injection aliquots and the pronuclei microinjection were performed by Dr. Oskar Ortiz from the IDG of the Helmholtz Center Munich. To prepare the sgRNAs and the Cas9 mRNA for the microinjection, injection aliquots were made as follows: single-use aliquots (30 µL) contained 25 ng/µL Cas9 mRNA, 12.5 ng/µL sgRNA mCRISPR1 and 12.5 ng/µL sgRNA mCRISPR3 in T<sub>10</sub>E<sub>0.1</sub> injection buffer. A master mix of 150 µL was prepared for up to 5 microinjections. This master mix was then filtered using a centrifugal filter and stored in 30 µL aliquots at -80 °C.

These microinjection aliquots were injected into the pronuclei (male) of one-cell embryos from female C57BL/6 wt<sup>182,183</sup> and cultured in 50 µM SCR7 (Ligase IV inhibitor) containing KSOM medium at 37 °C and 5 % CO<sub>2</sub> in a humidified incubator, overnight. On the next day, the two-cell embryos were washed in KSOM medium and were transferred into foster mothers

## 3.4 Genotyping of mouse lines

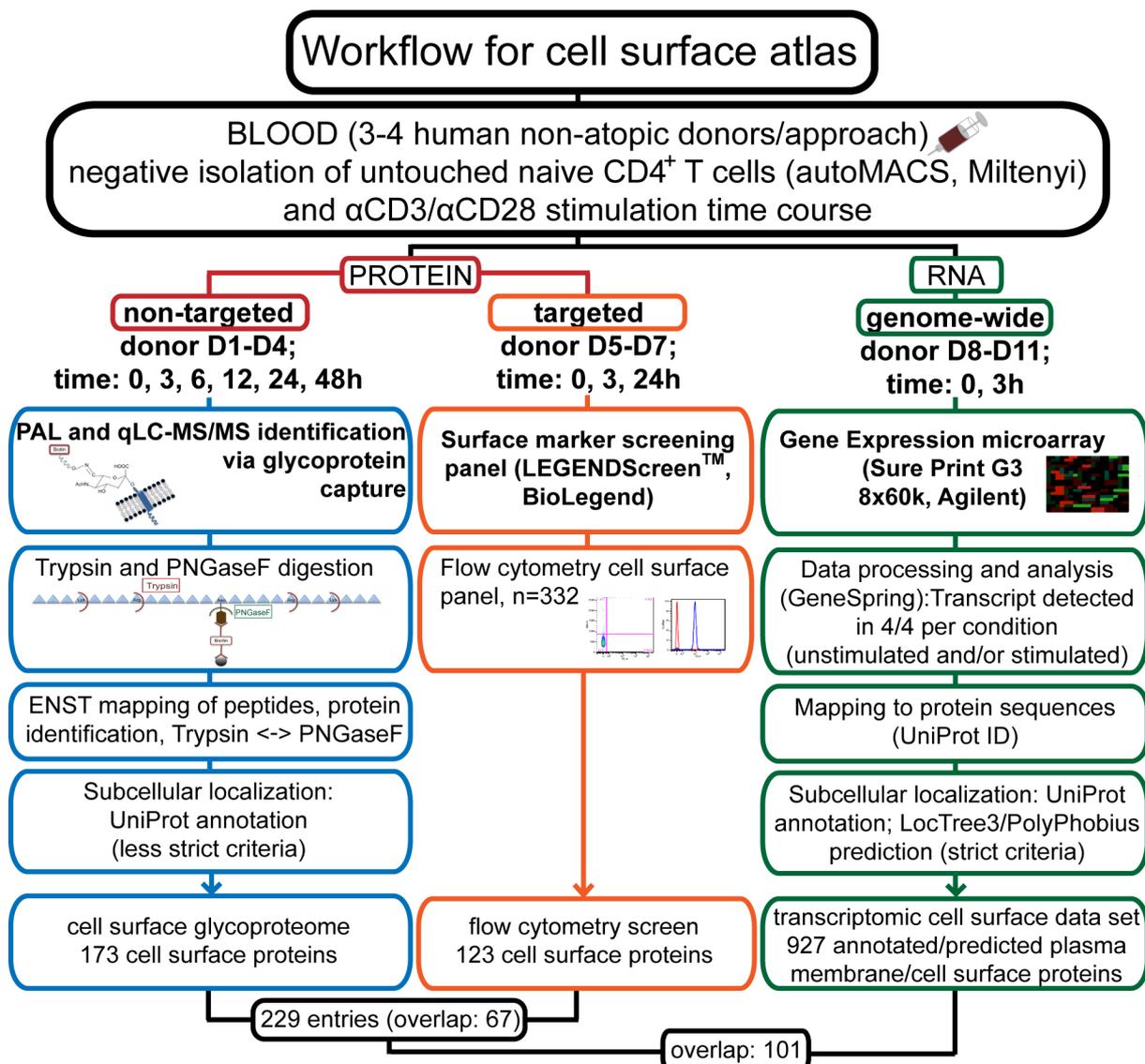
To genotype mouse lines, tail biopsies or biopsies from ear clippings were incubated in 50 µL biopsy lysis buffer at 96 °C for 10min. Samples were briefly vortexed and cooled down to RT. Another 50 µL biopsy lysis buffer and 0.5 µg Proteinase K were added and incubated at 56 °C

---

overnight. On the next day, the samples were vortexed and heat-inactivated at 96 °C for 15min. The samples were centrifuged at 6000 x g for 10min and stored at -20 °C. 1 µL of the supernatant was used for the genotyping PCR, which was carried out with 12.5 µL EconoTaq PLUS 2 x Master Mix and 10 µM forward and reverse primer, in a total volume of 25 µL. The PCR protocol was programmed with the following conditions: 95 °C (3min); [94 °C (30sec), 56 °C (40sec), 72 °C (30-60sec)] x 30 cycles; 72 °C (5min), 4 °C hold. The elongation time (1 min/kb) was adjusted to the amplicon size of each pair of primers, which can be found in Material Table M 5.

## VI. Results

### 1. Generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells



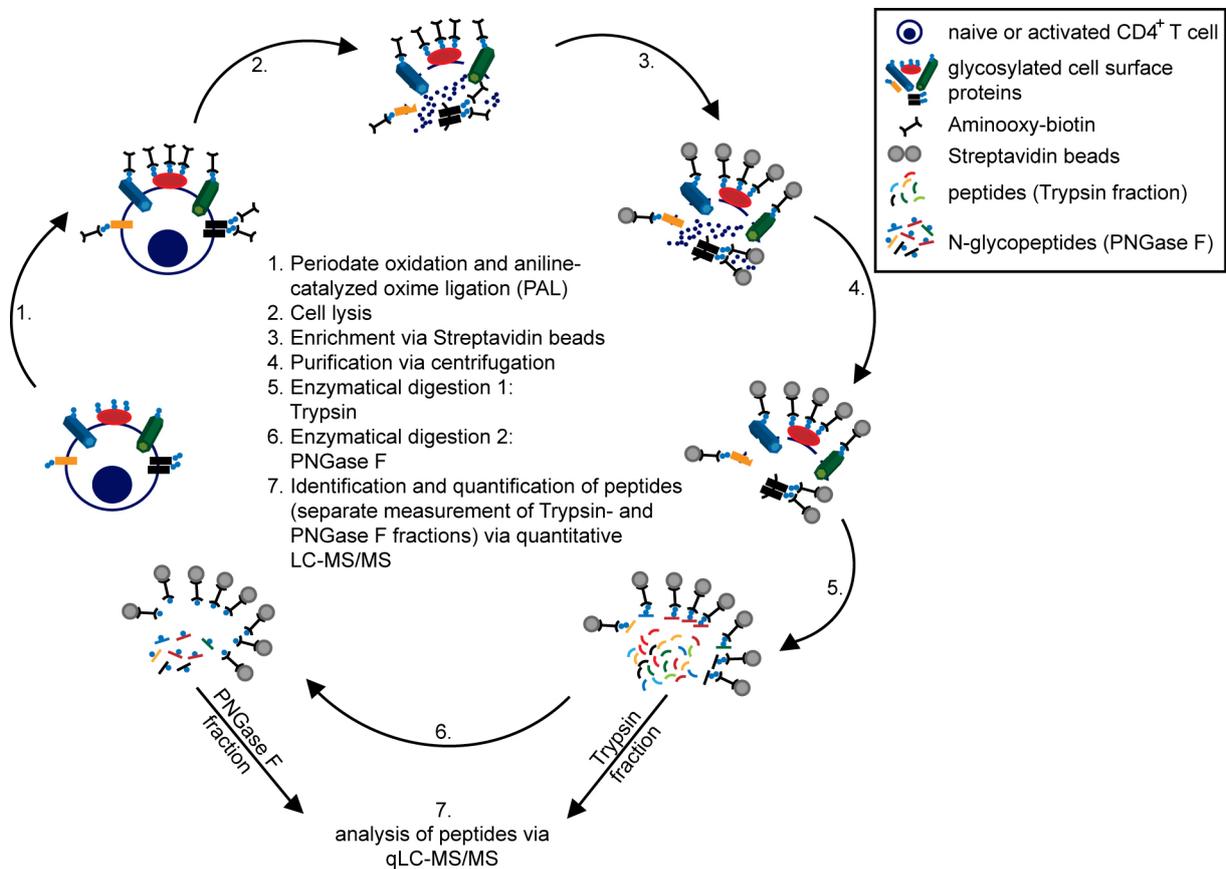
**Figure 4: Workflow for the generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells.** This overview shows how samples were generated and processed for two different proteomic techniques (PAL-qLC-MS/MS and flow cytometry screen) and one transcriptomic approach (genome-wide gene expression microarray analysis). The combination of the results of these techniques led to an extensive cell surface atlas for human naive and activated CD4<sup>+</sup> T cells. Starting with the isolation of untouched naive CD4<sup>+</sup> T cells from 3-4 human non-atopic blood donors per technique track (blue, orange, green) and their activation via anti-CD3/anti-CD28 stimulation in a time course experiment to mimic T cell receptor engagement. The processing steps for the different techniques and data analysis steps are given in the workflow, as well as the resulting number of identified cell surface proteins per approach. The combined entries of the two proteomic approaches, as well as the overlap between the different techniques are shown in the black boxes at the end of the workflow. (D=donor) (modified after <sup>157</sup>)

The generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells, which is described in the presented study, is based on a systematic multi-omic level workflow. This workflow focuses on two proteomic approaches, but also complements this level with one transcriptomic approach (Fig. 4).

At the beginning of each of the different techniques (Fig. 4: blue, orange, green), the workflow (Fig. 4) started with the isolation of untouched naive CD4<sup>+</sup> T cells from 3-4 human non-atopic blood donors (per technique) by magnetic separation. These isolated cells were either used in their naive state or stimulated with anti-CD3/anti-CD28 in a time course dependent manner to mimic T cell receptor engagement during an early window of T cell activation (0-48h). The first proteomic technique is called PAL-qLC-MS/MS (periodate oxidation and aniline-catalyzed oxime ligation coupled to quantitative liquid chromatography-tandem mass spectrometry) (Fig. 4, blue track) and achieves the idea of non-targeted and label-free identification and quantification of cell surface proteins. It is based on the fact that most cell surface proteins are glycosylated. Via this technique it was possible to specifically label, enrich and purify glycosylated cell surface proteins, which were subsequently identified and quantified via mass spectrometry. The second proteomic approach (Fig. 4, orange track) was a targeted high-throughput flow cytometry screen, which created the ability to test 332 monoclonal antibodies. On the transcriptomic level, a genome-wide gene expression microarray coupled to bioinformatics analysis was applied. The results were all combined in the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells (Table 2 and Fig. 20).

### **1.1 Establishment, validation and technical monitoring of PAL-qLC-MS/MS technique**

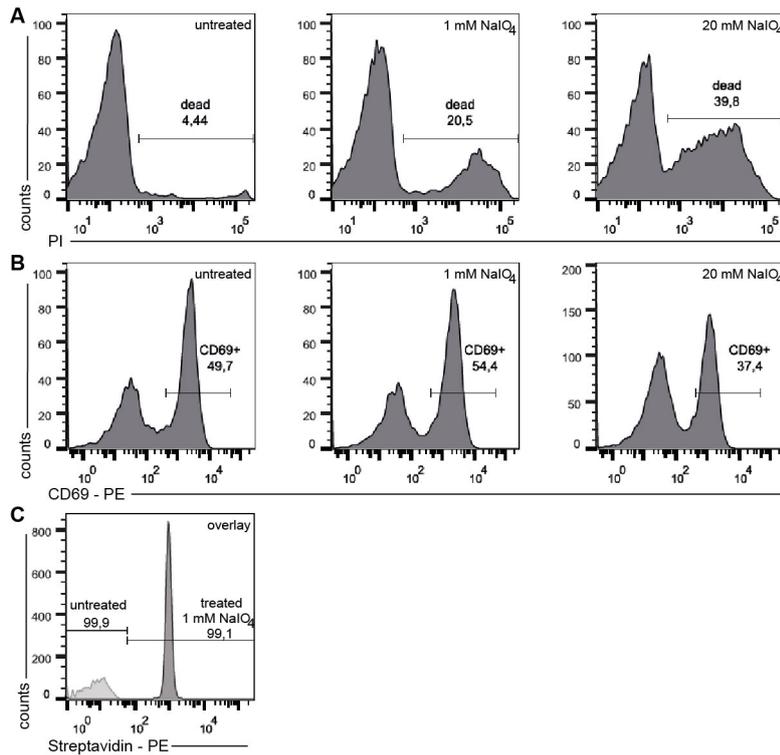
The aim of this study was to adapt the previously described PAL (periodate oxidation and aniline-catalyzed oxime ligation) technique for cell surface protein labeling and enrichment<sup>156</sup> for the use with primary human T cells and to prepare the samples suitable for mass spectrometry analysis. The PAL technique is based on the fact that most cell surface proteins are glycosylated.<sup>184</sup> Periodate is used to oxidize the alcohol groups of the sugar residues to form aldehydes. Aniline is then catalyzing the reaction in which the aldehyde forms a stable oxime-linkage to the Biotin-derivate Aminoxy-Biotin (Fig. 3). After this reaction, the glycosylated cell surface proteins are stably labeled with Aminoxy-Biotin, the cells are lysed and frozen. To then identify the Biotin-tagged cell surface proteins, the PAL-technique was complemented with quantitative liquid chromatography-tandem mass spectrometry (qLC-MS/MS). Therefore, the tagged cell surface proteins were purified and enriched via streptavidin beads and the proteins were digested first with Trypsin, followed by a PNGase F digest. The two enzymatically digested peptide fractions were separately kept and measured via qLC-MS/MS (Fig. 5).



**Figure 5: Scheme of the PAL-qLC-MS/MS technique.** This overview presents the steps of the PAL-qLC-MS/MS technique. The cells are oxidized and biotinylated via PAL technique (1.) and then lysed (2.). The biotinylated cell surface proteins are enriched via Streptavidin beads (3.) and purified via centrifugation (4.). The proteins are first enzymatically digested with Trypsin (5.) and the digested peptides are separated as Trypsin fraction. The Streptavidin beads coupled to the remaining peptides are incubated with PNGase F as a second enzyme for digestion (6.) and the resulting peptides separated as PNGase F fraction. Both fractions are analyzed separately via quantitative liquid chromatography-tandem mass spectrometry (qLC-MS/MS) (7.).

### 1.1.1 Influence of oxidation and biotinylation process

Oxidation agents like  $\text{NaIO}_4$  are known to be critical for the survival of cells, therefore the influence of different concentrations of  $\text{NaIO}_4$  was tested in terms of cell survival, detection of protein expression and biotinylation efficiency as an adequate labeling technique (Fig. 6).

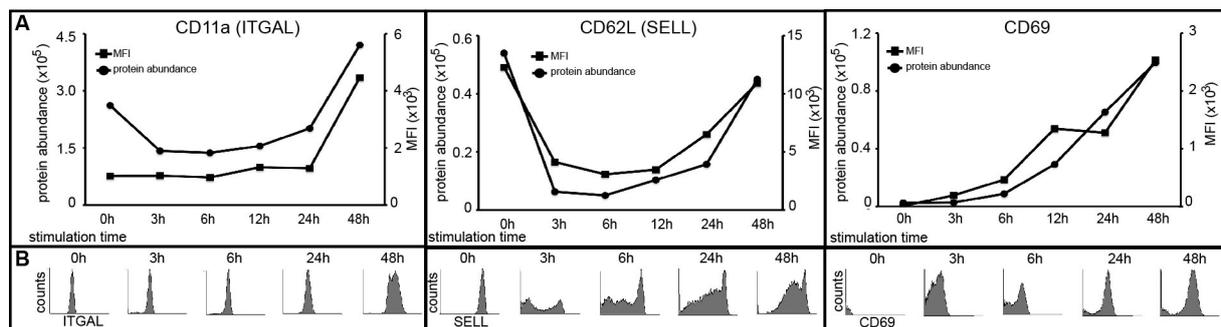


**Figure 6: Establishment and validation of PAL-qLC-MS/MS.** Different  $\text{NaIO}_4$  concentrations in the one-pot reaction of the PAL-qLC-MS/MS as well as the Biotin labeling efficiency were tested on living human  $\text{CD4}^+$  T cells, which were activated for 17h with anti-CD3/anti-CD28, and analyzed via flow cytometry. A) Staining of activated  $\text{CD4}^+$  T cells with PI to check cell viability upon oxidation treatment (1 mM or 20 mM  $\text{NaIO}_4$ ). B) Staining of activated  $\text{CD4}^+$  T cells with anti-CD69 antibody upon oxidation treatment (1 mM or 20 mM  $\text{NaIO}_4$ ). C) Biotinylation efficiency of cell surface proteins, which were either untreated or treated with  $\text{NaIO}_4$  was checked via Streptavidin-PE staining. The graph presents an overlay of two experiments showing the Streptavidin staining of untreated and treated cells.

1mM and 20 mM of  $\text{NaIO}_4$  were tested within the oxidation/biotinylation mix. Naive  $\text{CD4}^+$  T cells from one human blood donor were activated with anti-CD3/anti-CD28 for 17h.  $8 \times 10^6$  cells each were incubated without  $\text{NaIO}_4$ , with the 1mM or the 20 mM  $\text{NaIO}_4$  containing oxidation/biotinylation one-pot mix and stained for flow cytometry analysis. Fig. 6A shows that a concentration of 20 mM  $\text{NaIO}_4$  decreased the viability of cells about 39.8 % compared to 1 mM  $\text{NaIO}_4$  treatment, where 20.5 % of the cells are PI positive (untreated cells exhibit around 4.4 % of dead cells). The staining for CD69 was performed as a control, because CD69 is a well characterized T cell activation marker, and this showed that a concentration of 20 mM  $\text{NaIO}_4$  decreases the CD69 expression about 15 % (Fig 6B). These results pointed to a working concentration of 1 mM  $\text{NaIO}_4$ , but high biotinylation efficiency still needed to be ensured. A staining of cells with Streptavidin-PE, which were treated with 1 mM  $\text{NaIO}_4$  containing oxidation/biotinylation mix, confirmed a sufficient biotinylation efficiency of 99 % (Fig. 6C).

### 1.1.2 Validation of protein expression via flow cytometry in parallel to PAL-qLC-MS/MS sample preparation

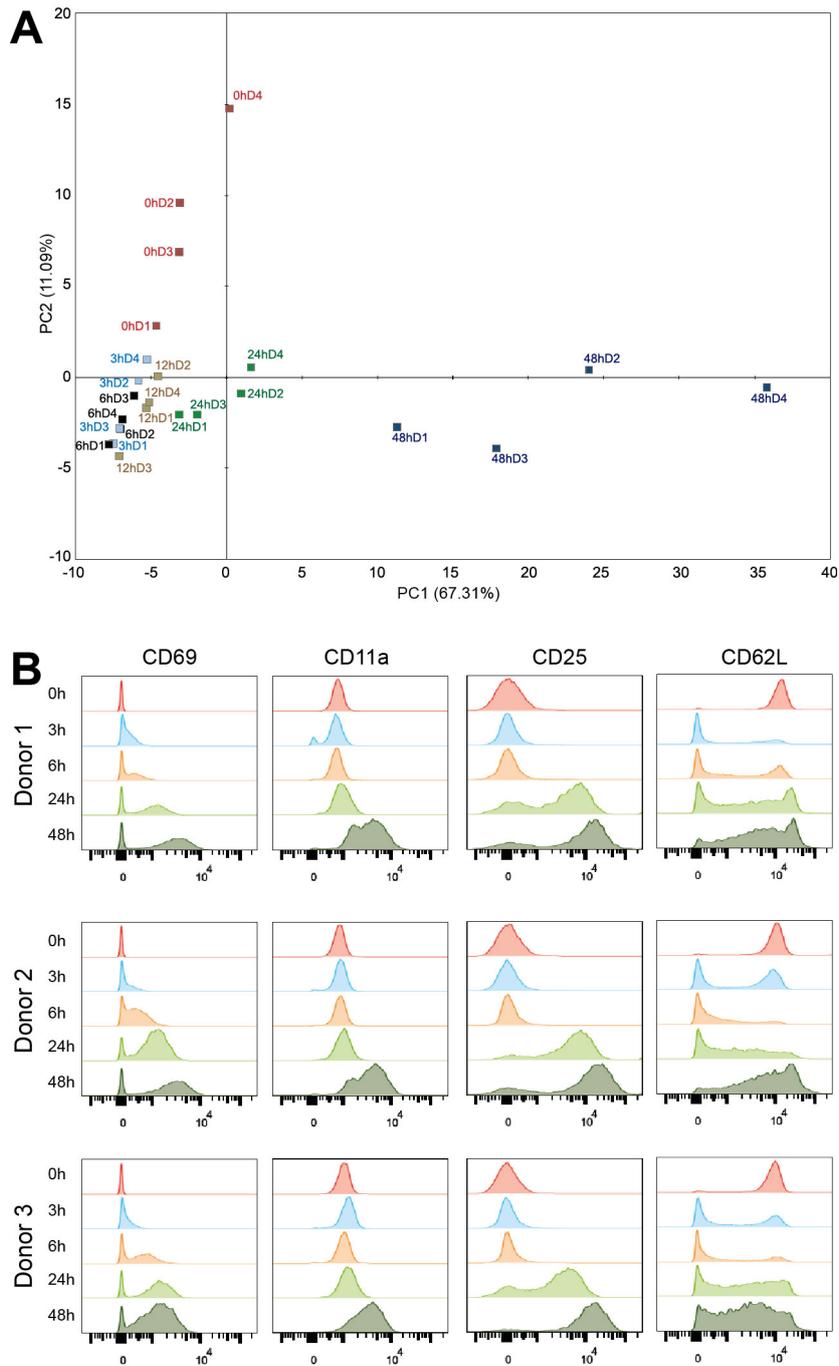
To ensure that the results of the quantitative protein expression measurements via PAL-qLC-MS/MS were comparable to the outcome of another validated and widely used technique, the samples for the generation of the surface glycoproteome (D1-D4) were stained for flow cytometry analysis, in parallel to sample processing for PAL-qLC-MS/MS. As targets for this validation staining the known T cell markers CD11a, CD62L and CD69 were chosen, because they already appeared in the mass spectrometry results during the establishment phase of the PAL-qLC-MS/MS technique. Fig. 7A shows that the expression pattern of the three proteins during the T cell activation obtained via MS (protein abundance) equals to the pattern measured via flow cytometry (MFI) (shown for one representative donor). Only a slight difference between the results of the techniques can be seen for CD69. The expression change between 12h and 24h of stimulation detected by flow cytometry showed a constant increase in contrast to the MS result, which described more a static expression state of the proteins between 12 and 24h of activation.



**Figure 7: Technical monitoring of T cell marker expression during sample preparation for PAL-qLC-MS/MS via flow cytometry.** For the validation of the PAL-qLC-MS/MS technique, a flow cytometry staining of the surface antigens CD11a, CD62L and CD69 was performed in parallel to sample preparation for PAL-qLC-MS/MS with T cells of the same donor. A) Protein abundance (PAL-qLC-MS/MS) values and mean fluorescence intensities (MFI; flow cytometry) are shown for the respective cell surface proteins at the respective time points (one representative donor). The expression pattern obtained via both techniques over the time course is comparable to each other. B) Histograms of the fluorescence intensity obtained by flow cytometry for the staining of the respective proteins at the respective time points are shown (one representative donor). (modified after <sup>157</sup>)

### 1.1.3 Assessment of donor variability by comparing the protein expression patterns

To be able to combine the results of the PAL-qLC-MS/MS technique of the four single different blood donors, expression patterns needed to be checked for similarity during the activation process. Therefore, the measured protein abundances of the different blood donors at the different stimulation time points were subjected to a principal component analysis and revealed highly concordant protein abundances (Fig. 8A).



**Figure 8: Comparability of the donor samples for PAL-qLC-MS/MS.** A) Protein abundances of the samples of the four donors (D1-D4) at all time points (0, 3, 6, 12, 24, 48h), obtained by PAL-qLC-MS/MS, were subjected to a Principal Component Analysis (PCA). This analysis grouped the donor samples at the time points (0h, 3-6-12h, 24h, 48h). Principal Component 1 (PC1), explaining 67.31 % of the data variance, divides all 48h samples from the other time points and PC2, explaining 11.09 % of the variance, separates the 0h samples from the samples of the remaining time points.<sup>157</sup> B) The donor comparability was also assessed via flow cytometry staining during the stimulation time course for the surface antigens CD69, CD11a, CD25 and CD62L for three donors (D1-D3) at the indicated time points (0, 3, 6, 24, 48h), demonstrating that the obtained expression patterns for the selected surface antigens are comparable between the different donors.

For selected T cell surface markers (CD69, CD11a, CD25, CD62L) the comparability of three of the donors was in addition assessed via flow cytometry (Fig. 8B), which also proved the

possibility to combine the datasets obtained for the different blood donors. Fig. 8B shows a continuous increase of the CD69 signal for the depicted donors, being able to distinguish a positive and a negative population at 24h. The CD11a expression was low until the increase between 24 and 48h, as well as the CD25 expression until 24h, also shown for the three donors. CD62L was highly expressed on the naive T cells of all examined donors, rapidly down-regulated after the start of the activation process, but increased again around the 24h time point. The results of the flow cytometry staining also showed high concordance between the donors included in the surface glycoproteome.

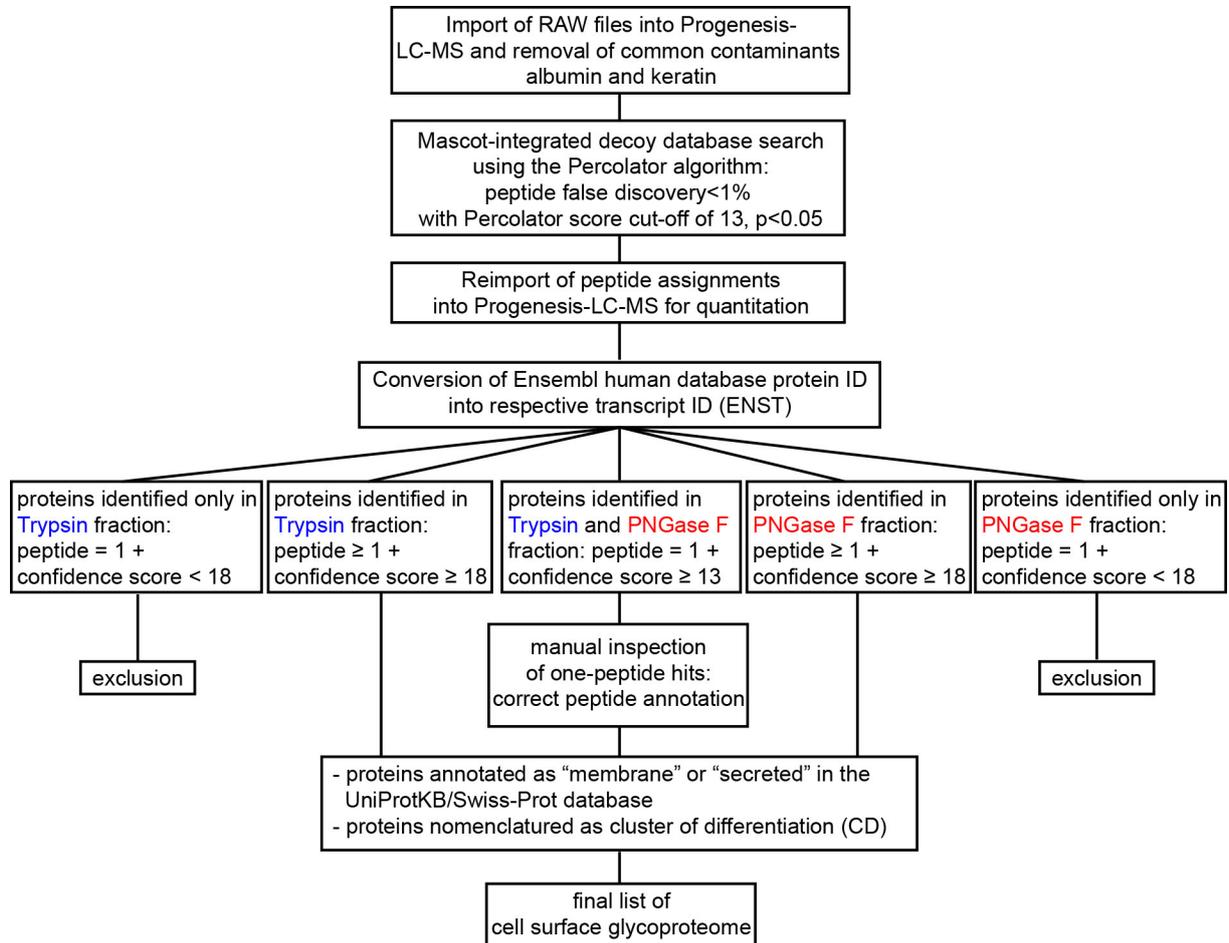
## **1.2 PAL-qLC-MS/MS-based cell surface glycoproteome of human naive and activated CD4<sup>+</sup> T cells**

To investigate changes in protein expression of the surface composition of human naive CD4<sup>+</sup> T cells, these cells were isolated from four non-atopic donors and taken in their naive state or stimulated with anti-CD3/anti-CD28 in a time course experiment for 3, 6, 12, 24 and 48h. The samples were prepared and measured via the PAL-qLC-MS/MS technique (Fig. 4, blue track; Fig. 5) and in parallel the technique was validated via flow cytometry (section 1.1.2) as well as the donor comparability was assessed (section 1.1.3). To set up the list of detected and verified cell surface proteins for the cell surface glycoproteome, the results of the mass spectrometry measurements, meaning the list of identified peptides converted to the corresponding proteins, were strategically evaluated.<sup>157</sup>

### **1.2.1 Strategic evaluation of raw qLC-MS/MS results and implementation of Trypsin and PNGase F fractions into the cell surface glycoproteome**

The RAW files (Thermo Fisher Scientific) of the qLC-MS/MS measurement were analyzed with Progenesis LC-MS software and the sums of the normalized abundances of all unique peptides were assigned to the respective proteins (Supplemental Table S1 and S2). To calculate the average peptide false discovery rate, a Mascot-integrated decoy database search combined with the use of the calculated Percolator algorithm was conducted. A peptide false discovery rate of < 1 % was calculated by performing the searches with a Percolator score cut-off of 13 and setting the significance threshold to  $p < 0.05$ . The search results and spectral files have been uploaded to the ProteomeXchange platform (<http://www.proteomexchange.org>) and are available with the identifier PXD001432. After removal of the common contaminants albumin and keratin, all Ensembl human database protein IDs were converted into the respective transcript IDs (ENST), aspiring highest possible stringency. Fig. 9 depicts the workflow how proteins from Trypsin and PNGase F

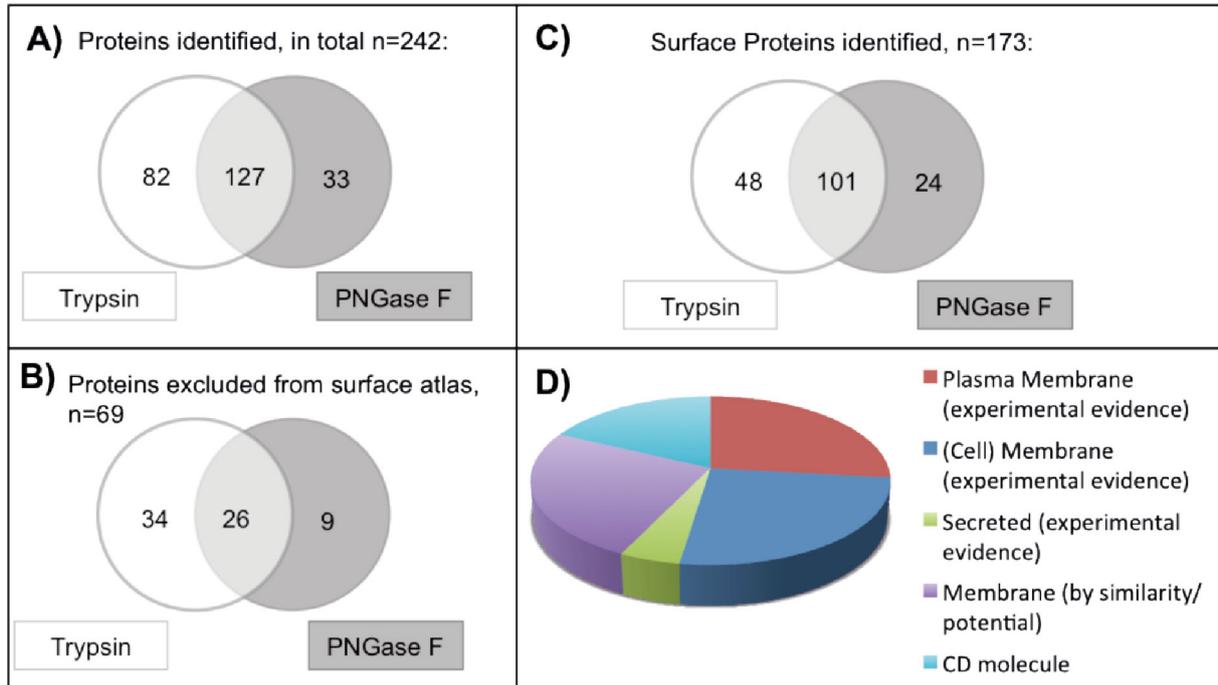
fractions were further taken into account for or removed from further analysis to generate a high confident dataset. For this decision tree the number of identified peptides, the identification in one or both enzymatic digestion fractions (Trypsin and/or PNGase F), and the assigned confidence score are the critical points for each protein.



**Figure 9: Decision Tree for the analysis of qLC-MS/MS-derived data to assemble the cell surface glycoproteome.** This decision tree was built to analyze and evaluate the RAW files (Thermo Fisher Scientific), which were derived from the qLC-MS/MS measurement, to be able to create the list of detected cell surface proteins with high confidence.

Fig. 10 summarizes the results after evaluating the protein lists from the different enzymatic digestion fractions, according to the decision tree of Fig. 9. In total 242 proteins were identified by PAL-qLC-MS/MS on naive and/or activated CD4<sup>+</sup> T cells (Fig. 10A). Based on the decision tree (Fig. 9), 69 proteins were not included into the cell surface glycoproteome (Fig. 10B). 173 of the identified proteins (Fig. 10C) were included in the surface glycoproteome, based on their subcellular localization (UniProt\_SL) annotation, which was given by the UniProtKB/Swiss-Prot database. The annotation as CD molecule or the UniProt\_SL annotation as "plasma membrane", "cell membrane" or "secreted" with experimental evidence was already given for 131 of the 173 cell surface proteins. This surface glycoproteome dataset expands the so far existing knowledge of plasma membrane-associated localization for the

remaining 42 proteins by providing experimental evidence due to the PAL-qLC-MS/MS approach (Fig. 10D).<sup>157</sup>



**Figure 10: Illustration of the number of identified proteins of the PAL-qLC-MS/MS approach.** A) Venn diagram showing the number of identified proteins in the Trypsin and PNGase F fraction, as well as the number of proteins overlapping both fractions. B) Venn diagram displaying the number of proteins from Trypsin and/or PNGase F fractions, which were excluded from the cell surface atlas. C) Venn diagram showing the number of cell surface proteins, which were identified in one or both of the fractions. D) Pie chart illustrating the subcellular localization distribution of the identified cell surface proteins, which is given by the UniProtKB/Swiss-Prot database (given as experimental evidence or by similarity to closely related homologs or defined as potential).<sup>157</sup>

A complete list of the 173 cell surface proteins, which were identified on naive and/or activated CD4<sup>+</sup> T cells by PAL-q-LC-MS/MS, is given in Table 2 (marked by a triangle under “identified via”). This list also provides the information about an alternative name for the proteins, the respective transcript ID (ENST), the UniProt ID, and the number of peptide counts. In addition, Table 2 shows if the protein is expressed or is not expressed (letters in bold) on the naive CD4<sup>+</sup> T cell.

CD name	Description	ENST	UniprotID	peptide counts	identified via	DrugBank
CD2	LFA-2	ENST00000369478	P06729	8	▲●	a
CD3D	T3D	ENST00000300692	P04234	3	▲	a
CD3G (1)	T3G	ENST00000392883	A8MUH3	2	▲●	a
CD3G (2)		ENST00000532917	P09693	2	▲●	a
CD4	T4	ENST0000011653	P01730	6	▲●	
CD5	LEU1	ENST00000347785	P06127	17	▲●	
CD6	T12	ENST00000313421	P30203	9	▲●	
CD7	gp40	ENST00000312648	P09564	4	▲●	
CD9	MIC3		P21926		▲●	i
CD11a	ITGAL	ENST00000356798	P20701	12	▲●	a
CD15	FUT4		P22083		▲●	a
CD16	FCGR3A		P08637		▲●	a
CD18	ITGB2	ENST00000302347	P05107	17	▲●	a
CD25	IL2RA	ENST00000256876	P01589	1	▲●	a
CD26	DPP4	ENST00000360534	P27487	5	▲●	a
CD27	TNFRSF7	ENST00000266557	P26842	4	▲●	
CD28	TP44	ENST00000324106	P10747	3	▲●	
CD29	ITGB1	ENST00000302278	P05556	3	▲●	a
CD31	PECAM1	ENST00000563924	P16284	6	▲●	
CD37	TSPAN26	ENST00000323906	P11049	1	▲●	
CD38	ADP-ribosyl cyclase	ENST00000226279	P28907	4	▲●	
CD43	SPN	ENST00000360121	P16150	4	▲●	
CD44	LHR	ENST00000263398	P16070-12	13	▲●	a
CD45	PTPRC	ENST00000352140	E9PC28	58	▲●	
CD45RA	PTPRC				▲●	
CD45RB	PTPRC				▲●	
CD45RO					▲●	
CD46 (1)	MCP	ENST00000322875	P15529-2	5	▲●	
CD46 (2)		ENST00000358170	P15529	3	▲●	i
CD47	MER6	ENST00000355354	Q08722-3	3	▲●	
CD48	BCM1	ENST00000368046	P09326	21	▲●	
CD49d	ITGA4	ENST00000233573	P13612	1	▲●	a
CD49e	ITGA5	ENST00000293379	P08648	1	▲●	
CD49f	ITGA6		P23229		▲●	
CD50	ICAM3	ENST00000160262	P32942	6	▲●	
CD52	CAMPATH-1		P31358		▲●	a
CD53	MOX44		P19397		▲●	
CD54	ICAM1	ENST00000264832	P05362	10	▲●	a
CD55	DAF	ENST00000314754	P08174	2	▲●	a
CD58	LFA-3	ENST00000369487	P19256	1	▲●	
CD59	Protectin	ENST00000351554	P13987	3	▲●	
CD61	ITGB3		P05106		▲●	a
CD62L	SELL	ENST00000236147	P14151-2	4	▲●	
CD63	MLA1		P08962		▲●	
CD69	CLEC2C	ENST00000228434	Q07108	5	▲●	
CD71	TFRC	ENST00000360110	P02786	18	▲●	a
CD73	NT5E		P21589		▲●	a
CD74	DHLAG	ENST00000009530	P04233	1	▲●	i
CD79b	IGB		P40259		▲●	
CD81	TAPA1		P60033		▲●	
CD82	KAI1	ENST00000227155	P27701	6	▲●	
CD83			Q01151		▲●	
CD84	SLAMF5	ENST00000311224	Q9UIB8	3	▲●	
CD85j	LILRB1		Q8NHL6		▲●	a
CD85k	LILRB4		Q8NHJ6		▲●	a
CD95	FAS (1)	ENST00000352159	H7C5Z8	1	▲●	
	FAS (2)	ENST00000355740	P25445	3	▲●	
CD96	T-cell surface protein tactile		P40200		▲●	
CD97		ENST00000242786	P48960	11	▲●	
CD98	SLC3A2	ENST00000338663	P08195-2	26	▲●	
CD98LC	SLC7A5	ENST00000261622	Q01650	2	▲●	a
CD99	MIC2	ENST00000381187	P14209-3	1	▲●	
CD100	SEMA4D	ENST00000356444	Q92854	8	▲●	
CD101	IGSF2		Q93033		▲●	i
CD102	ICAM2	ENST00000418105	P13598	6	▲●	
CD105	ENG	ENST00000344849	P17813-2	2	▲●	
CD107a	LAMP1	ENST00000332556	P11279	4	▲●	
CD108	SEMA7A	ENST00000261918	Q5326	2	▲●	
CD109	CPAMD7	ENST00000287097	Q6YHK3	2	▲●	
CD116	CSF2RA		P15509		▲●	a
CD120b	TNFRSF1B	ENST00000376259	P20333	2	▲●	a
CD126	IL6R		P08887		▲●	a
CD127	IL7Rα		P16871		▲●	
CD129	IL9R		Q01113		▲●	
CD131	CSF2RB		P32927		▲●	a
CD132	IL2RG	ENST00000374202	P31785	1	▲●	a
CD134	TNFRSF4	ENST00000379236	P43489	2	▲●	
CD137	TNFRSF9		Q07011		▲●	
CD138	SDC1		P18827		▲●	i
CD144	CDH5		P33151		▲●	a
CD147	BSG	ENST00000333511	P35613	15	▲●	
CD148	PTPRJ	ENST00000418331	Q12913	1	▲●	
CD150	SLAMF1	ENST00000235739	Q13291	4	▲●	
CD152	CTLA4		P16410		▲●	a
CD153	TNFSF8	ENST00000223795	P32971	2	▲●	
CD154	CD40LG	ENST00000370628	Q3LBU2	1	▲●	
CD156c	ADAM10	ENST00000260408	O14672	2	▲●	i
CD158a/h	KIR2DL1		Q6IST4		▲●	
CD158f	KIR2DL5A		Q8N109		▲●	i
CD162	SELPLG	ENST00000228463	Q14242-2	2	▲●	
CD164	MUC-24		Q04900		▲●	
CD165	SLC38A5	ENST00000317669	Q8WUX1	2	▲●	
CD166	ALCAM	ENST00000306107	Q13740	2	▲●	i
CD172g	SIRPG		Q9P1W8		▲●	
CD183	CXCR3		P49682		▲●	
CD184	CXCR4		P61073		▲●	a
CD196	CCR6		P51684		▲●	
CD197	CCR7		P32248		▲●	
CD200	MOX1	ENST00000315711	P41217-2	3	▲●	
CD200R	CD200R1		Q8TD46		▲●	
CD205	Ly75		O60449		▲●	i
CD218a	IL18R1	ENST00000233957	Q13478	2	▲●	
CD226	DNAM-1	ENST00000280200	Q15762	2	▲●	
CD229	LY9	ENST00000263285	Q9HGB7	1	▲●	
CD230	PRNP	ENST00000379440	P04156	1	▲●	a
CD235ab	GYPA		P02724		▲●	
CD236	GYPC	ENST00000259254	P04921	1	▲●	
CD243	ABCBI	ENST00000265724	P08183	1	▲●	a
CD245 (p220/240)	Dy12, Dy35		Q14207, Q99973		▲●	
CD257	TNFSF13B		Q9Y275		▲●	a
CD258	TNFSF14	ENST00000245912	O43557	1	▲●	

Proteomic approaches:  
 ▲● PAL-qLC-MS/MS  
 ●○ Flow cytometry surface screen (expressed/not expressed)  
 comment: ★verified in only one donor (flow cytometry) but either additionally identified via PAL-qLC-MS/MS or known in the context of T cell biology  
 no sign = no antibody against this antigen in screening included

Transcriptomic approach:  
 ■□ transcript annotated/predicted as plasma membrane or cell surface protein (yes/no)  
 comment: 1) annotated/predicted as secreted  
 2) annotated/predicted as located in other cell compartment  
 3) detected in <4 donor samples  
 4a) LocTree3: secreted, but TMH=0  
 4b) LocTree3: plasma membrane, but TMH=0  
 5) could not be mapped to UniProt AC  
 ▨ putative dataset or annotation as "membrane"  
 no sign = no probe for this transcript on the array

§ not yet mentioned in the context of the biology of naive CD4+ T cells before  
**bold: not expressed on naive CD4+ T cells**  
*italic: expression differences obtained via PAL-qLC-MS/MS and flow cytometry*

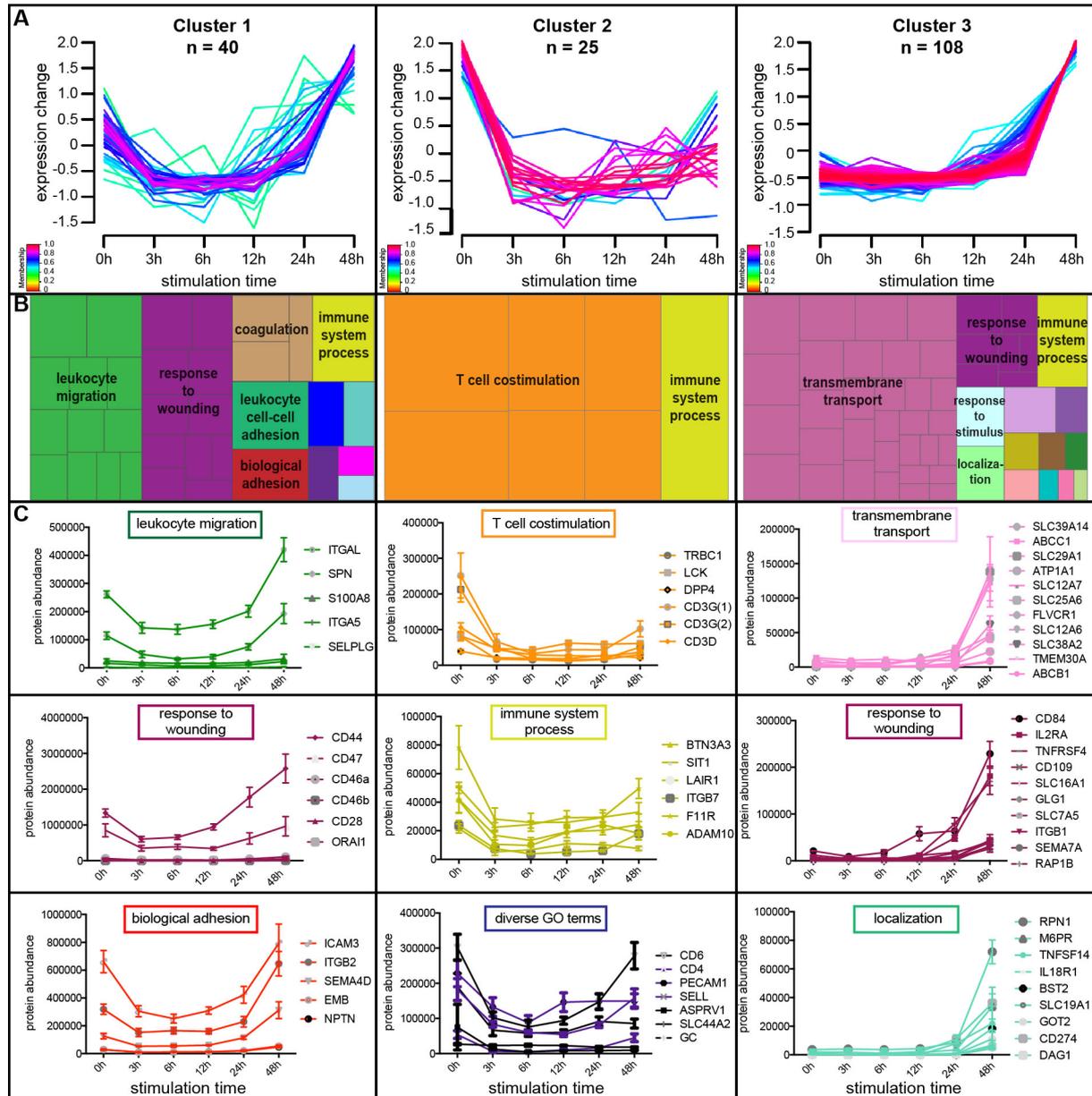
DrugBank: a=approved, e=experimental, i=investigational

CD270	TNFRSF14		Q92956		●	■
CD272	BTLA	ENST00000334529	Q726A9	1	▲	■
CD274	B7H1	ENST00000381573	Q9NZQ7-2	2	▲	■
CD277	BTN3A1		O00481		●	■
CD278	ICOS		Q9Y6W8		●	■
CD298	ATP1B3	ENST00000286371	P54709	9	▲	■
CD305	LAIR1	ENST00000348231	Q6GTx8	1	▲	■
CD315	PTGFRN	ENST00000393203	Q9P2B2	1	▲	■
CD316	IGSF8	ENST00000314485	Q969P0	4	▲	■
CD317	BST2	ENST00000252593	Q10589	3	▲	■
CD319	SLAMF7	ENST00000359331	Q9NQ25	1	▲	■
CD321	F11R	ENST00000289779	Q9Y624	2	▲	■
CD352	SLAMF6	ENST00000368057	Q96DU3	1	▲	■
CD357	TNFRSF18	ENST00000328596	Q9Y5U5	1	▲	■
CD360	IL21R		Q9HBE5		●	■
CD361	EVI2B	ENST00000330927	P34910	1	▲	■
	ABCC1	ENST00000345148	P33527-4	4	▲	■
	AMICA1	ENST00000292067	Q86YT9-2	3	▲	■
	APOC3 §	ENST00000227667	P02656	1	▲	■
	ASPRV1	ENST00000320256	Q53RT3	2	▲	■
	ATP1A1 §	ENST00000295598	P05023	2	▲	■
	B2M	ENST00000544417	P61769	2	▲	■
	BTN3A2	ENST00000356386	P78410	2	▲	■
	BTN3A3	ENST00000244519	O00478	4	▲	■
	C10orf54	ENST00000394957	Q9H7M9	2	▲	■
	C16orf54 §	ENST00000329410	Q6UWD8	2	▲	■
	CSL2, GPR77		Q9P296		●	■
	CALML5 §	ENST00000380332	Q9NZT1	2	▲	■
	CCR10		O00590		●	■
	CFI1	ENST00000308162	P23528	1	▲	■
	CORO1A	ENST00000219150	P31146	2	▲	■
	CPM	ENST00000338356	P14384	6	▲	■
	CXCR7		P25106		●	■
	DAG1	ENST00000308775	Q14118	1	▲	■
	DCD §	ENST00000293371	P81605	2	▲	■
	ECE1 §	ENST00000264205	P42892-4	5	▲	■
	EMB	ENST00000303221	Q6PCB8	2	▲	■
	ENO1	ENST00000234590	P06733	1	▲	■
	ERO1L §	ENST00000395686	Q96HE7	1	▲	■
	EVI2A §	ENST00000247270	P22794-2	1	▲	■
	FLVCR1	ENST00000366971	Q9Y5Y0	1	▲	■
	GC	ENST00000273951	P02774	2	▲	■
	GLG1	ENST00000205061	Q92896	3	▲	■
	GNAI2	ENST00000266027	P04899	2	▲	■
	GOT2 §	ENST00000245206	P00505	1	▲	■
	GPA33 §	ENST00000367868	Q99795	2	▲	■
	GPR171	ENST00000309180	O14626	4	▲	■
	HLA-A2		P01892		●	■
	HLA-A (1)	ENST00000376806	Q55RNS	15	▲	■
	HLA-A (2)	ENST00000438861		22	▲	■
	HLA-B (1)	ENST00000412585	P01889	11	▲	■
	HLA-B (2)	ENST00000435618		10	▲	■
	HLA-B (3)	ENST00000396935		14	▲	■
	HLA-C (1)	ENST00000457903	P04222	17	▲	■
	HLA-C (2)	ENST00000383323		13	▲	■
	HLA-E	ENST00000376630	P13747	7	▲	■
	HLA-G	ENST00000380323	P17893	2	▲	■
	HYOU1	ENST00000404233	Q9Y4L1	2	▲	■
	IL27RA	ENST00000263379	Q6UWB1	4	▲	■
	INHBB §	ENST00000295228	P09529	1	▲	■
	ITGB7	ENST00000267082	P26010	3	▲	■
	LCK	ENST00000333070	P06239	3	▲	■
	LDLR	ENST00000252444	J3KMZ9	2	▲	■
	LNPEP	ENST00000231368	Q9UIQ6	1	▲	■
	LTB	ENST00000376117	Q08643	2	▲	■
	M6PR	ENST00000000412	P20645	2	▲	■
	MPZL1	ENST00000359523	Q95297	1	▲	■
	MPZL2	ENST00000278937	O60487	1	▲	■
	MSC and NPC				●	■
	Notch 1		P46531		●	■
	Notch 2		Q04721		●	■
	NPTN §	ENST00000287226	Q9Y639	3	▲	■
	ORAI1	ENST00000330079	Q96D31	1	▲	■
	Podocalyxin, TRA-1-81		O00592		●	■
	PTPRA	ENST00000216877	P18433	2	▲	■
	PTPRCAP	ENST00000326294	Q14761	1	▲	■
	RAP1B	ENST00000250559	P61224	1	▲	■
	RNF149 §	ENST00000295317	Q8NC42	1	▲	■
	RPN1	ENST00000296255	P04843	2	▲	■
	S100A8	ENST00000368732	P05109	1	▲	■
	S1PR4	ENST00000246115	Q95977	1	▲	■
	SBSN §	ENST00000452271	Q6UWP8	2	▲	■
	SIT1	ENST00000259608	Q9Y3P8	3	▲	■
	SLC12A2	ENST00000262461	P55011	2	▲	■
	SLC12A6	ENST00000290209	Q9UHW9-2	2	▲	■
	SLC12A7	ENST00000264930	Q9Y666	1	▲	■
	SLC16A1	ENST00000369626	P53985	1	▲	■
	SLC19A1	ENST00000311124	P41440	1	▲	■
	SLC1A4 §	ENST00000234256	P43007	1	▲	■
	SLC1A5	ENST00000542575	Q15758	9	▲	■
	SLC20A1	ENST00000272542	Q8WUM9	1	▲	■
	SLC25A3 §	ENST00000188376	Q00325	1	▲	■
	SLC25A5 §	ENST00000317881	P05141	1	▲	■
	SLC25A6	ENST00000381401	P12236	1	▲	■
	SLC29A1	ENST00000313248	Q99808	6	▲	■
	SLC2A1	ENST00000426263	P11166	5	▲	■
	SLC2A3	ENST00000075120	P11169	3	▲	■
	SLC38A1	ENST00000398637	Q9H2H9	1	▲	■
	SLC38A2	ENST00000256689	Q96QD8	2	▲	■
	SLC39A14	ENST00000240095	Q15043-2	2	▲	■
	SLC43A3	ENST00000352187	Q8NB15	4	▲	■
	SLC44A2	ENST00000335757	Q8IWA5	5	▲	■
	SLC4A2	ENST00000392826	P04920-3	4	▲	■
	SLC4A7 §	ENST00000454389	Q9Y6M7-7	7	▲	■
	SLC5A3 §	ENST00000381151	P53794	2	▲	■
	SLC5A6 §	ENST00000310574	Q9Y289	2	▲	■
	SLC6A6	ENST00000360861	P31641	1	▲	■
	SLC7A1 §	ENST00000380752	P30825	3	▲	■
	SMR3B §	ENST00000304915	P02814	1	▲	■
	SYPL1 §	ENST00000011473	Q16563	2	▲	■
	TCR Vβ8				●	■
	TCR Vβ9				●	■
	TCR αβ				●	■
	TF	ENST00000402696	P02787	27	▲	■
	TGFBFR2	ENST00000295754	P37173	2	▲	■
	TMEM2 §	ENST00000377044	Q9UJN6	1	▲	■
	TMEM30A	ENST00000230461	Q9NV96	1	▲	■
	TRBC1	ENST00000570538	P01850	3	▲	■

**Table 2: List of all cell surface proteins identified on human naive and/or activated CD4<sup>+</sup> T cells.** This table includes all cell surface proteins detected on protein level on human naive and/or activated CD4<sup>+</sup> T cells, flagged by the technique, which identified the respective protein. As far as possible the Ensembl human database transcript ID and the UniProtKB ID are given for the respective protein. If the protein was identified via PAL-qL-MS/MS the number of identified peptides (peptide count), is listed. In addition the information about the status of the protein as target of a drug is given (DrugBank).<sup>157</sup>

### 1.2.2 Analysis of cell surface protein expression patterns by unsupervised clustering and Gene Ontology enrichment analysis of resulting expression clusters

To investigate if the 173 proteins (detected by PAL-qLC-MS/MS) exhibit different expression patterns during the stimulation time course of the naive CD4<sup>+</sup> T cell, an unsupervised expression cluster analysis (Fig. 11A) was performed.<sup>157</sup> The dynamic changes of the protein abundances during stimulation could be grouped into three clusters (Supplemental Table S4). Cluster 1 (40 proteins) and Cluster 2 (25 proteins) show a fast decrease in expression already during the first hours of activation. The difference between Cluster 1 and 2 was seen during the later stimulation time points. The expression pattern of Cluster 1 is marked by a later increase after the first decrease and the expression level of proteins contained in Cluster 2 stayed low and did not reach the initial expression level seen on the naive CD4<sup>+</sup> T cells. The largest cluster is Cluster 3 (108 proteins). The expression level of the proteins belonging to this cluster is in general low during the first stimulation time points and changes only marginally in the beginning, but around the 24h time point, the protein expression strongly increases. To characterize the proteins contained in the different clusters on another level than dynamic expression profiles, a Gene Ontology (GO) Enrichment Analysis was performed. Enriched GO terms, which were overlapping between the proteins contained in one cluster, were identified via the Generic GO TermFinder algorithm. The software REVIGO was then utilized to summarize the GO terms as “superclusters” and to visualize the obtained GO terms as treemaps (Fig. 11B, more details about the GO terms within the “superclusters” are given in Appendix Fig. A1-A3). Some GO terms, predominantly immune system related GO terms (e.g. immune system process, response to stimulus), were common between Cluster 1, 2 and 3. But there are also very specific GO terms, which are not shared among all clusters, like e.g. proteins of Cluster 1 are specifically involved in migration and adhesion. Proteins of Cluster 2 shared very specific T cell-related terms like T cell co-stimulation and activation and proteins of Cluster 3 are mainly involved in transmembrane transport. Fig. 11C shows the expression profiles of selected proteins belonging to the different clusters. Plotted in one graph and color coded according to the “superclusters” in the treemaps (Fig. 11B), are proteins which share the same GO term heading. In general, most proteins are assigned to more than one GO term, but the listed proteins were specifically selected to represent one specific GO term “supercluster”.

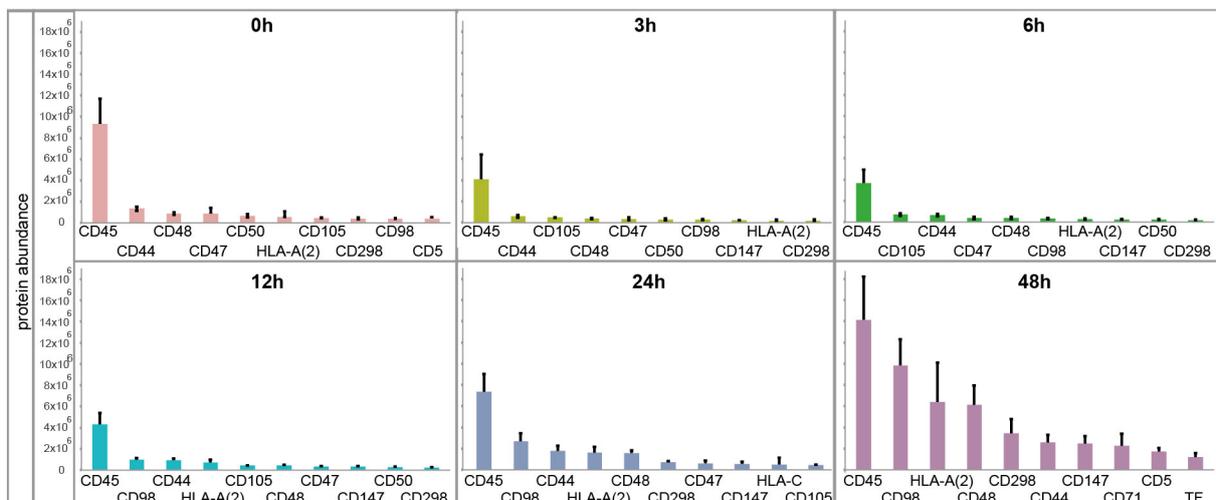


**Figure 11: Unsupervised clustering and Gene Ontology enrichment analysis of the cell surface proteins of human naive and activated CD4<sup>+</sup> T cells, identified by PAL-qLC-MS/MS.** A) GProX cluster analysis of 173 cell surface proteins (identified by PAL-qLC-MS/MS) according to their quantitative expression changes (normalized protein abundances) during the stimulation time course (0, 3, 6, 12, 24, 48h) with anti-CD3/anti-CD28. The color of each line, corresponding to one protein, represents the membership value of this protein, explaining how exact the expression of this protein fits to this cluster. Only proteins with a membership value above 0.4 are shown (a table including all membership values for the respective proteins in the different clusters is given in Supplemental Table S4). B) REVIGO treemaps was used to visualize GO (Gene Ontology) term analysis for the proteins belonging to the three different clusters with at least a membership value of 0.6. Each single rectangle is a representation of an enriched GO term and these GO terms are combined as “superclusters” of loosely connected terms, shown by the different colors. The p-value of the GO term in the underlying GOA database is visualized by the adjusted size of the rectangles.<sup>167</sup> A detailed figure including the GO terms of the smaller rectangles can be found in the Appendix Fig. A1-A3. C) Expression profiles of selected proteins during anti-CD3/anti-CD28 stimulation (0, 3, 6, 12, 24, 48h) (n=4, mean +/- SEM), grouped according to their allocated GO terms and flagged by the related rectangle color. (modified after<sup>157</sup>)

### 1.2.3 Differences and similarities between naive and activated CD4<sup>+</sup> T cells

To describe the differences between the naive and the anti-CD3/anti-CD28 activated CD4<sup>+</sup> T cell, the proteins, which were identified on naive and/or activated CD4<sup>+</sup> T cells, were compared. Table 2 provides the information about the presence or absence of a protein on the naive CD4<sup>+</sup> T cell in general. Due to the PAL-qLC-MS/MS results there are six proteins (CD98LC, CD120b, CD218a, CD258, CD272, CD357) which were not detected on naive CD4<sup>+</sup> T cells, meaning that they are expressed at later stimulation time points and are therefore no cell surface markers for human naive CD4<sup>+</sup> T cells. Whereas for the expression of further ten proteins (SBSN, DAG1, Fas-(2), HLA-B-(3), RNF149, S1PR4, SLC1A4, SLC6A6, TNFRSF18, TNFSF8), the detection by PAL-qLC-MS/MS was different between the naive CD4<sup>+</sup> T cells of the four different donors. But these ten proteins were more stably detected among the four donor samples during later time points of the T cell activation time course.

Fig. 12 provides an overview about the ten proteins, which had the highest measured protein abundance at every indicated stimulation time point. In general, this figure points out that the group of the highest expressed proteins is nearly the same at all time points, only a few proteins differ between the time points. CD45 is the highest abundant protein at every time point, starting from the naive CD4<sup>+</sup> T cell to the T cell, which was activated for 48h. CD44, CD48, HLA-A(2), CD298 and CD98 are also among the ten highest expressed cell surface proteins at all examined time points. This figure also shows the general level of measured protein abundance between the different time points. It shows that the measured level is comparable at all time points from 0 – 24h (highest value ranging from 3.8 to 9 x 10<sup>6</sup> units), only at 48h the protein abundance is considerable higher for nearly all of the ten highest expressed proteins (ten highest values ranging from 1.8 to 14 x 10<sup>6</sup> units).



**Figure 12: Cell surface proteins exhibiting the highest protein abundance (PAL-q-LC-MS/MS) at the depicted stimulation time points.** Overview of the 10 highest expressed cell surface proteins at every examined stimulation time point (0, 3, 6, 12, 24, 48h) (n=4, mean + SEM) during the stimulation time course experiment.

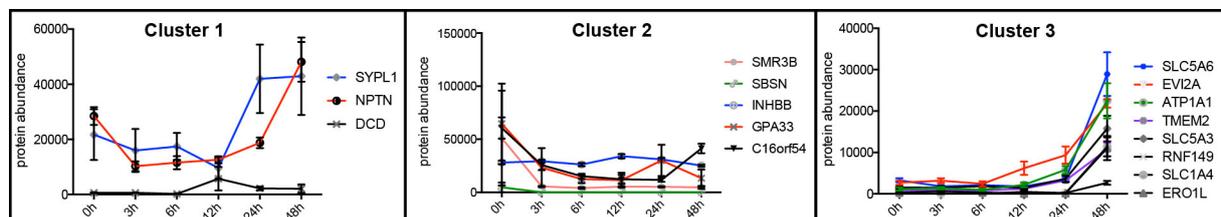
### 1.2.4 Identification and validation of cell surface proteins, which were not co-cited with T cell biology before

An extensive information search was performed to identify cell surface proteins, which were not mentioned in the context of T cell biology before.<sup>157</sup> Every cell surface protein (n=173) of the cell surface glycoproteome has undergone a closer examination by a NCBI PubMed literature search, a detailed UniProtKB/Swiss-Prot information- as well as a google.com/patent search. 86 % of the cell surface proteins were already co-cited within T cell biology or at least immune cell biology linked contexts, supporting the capability of the PAL-qLC-MS/MS technique to efficiently identify cell surface proteins.

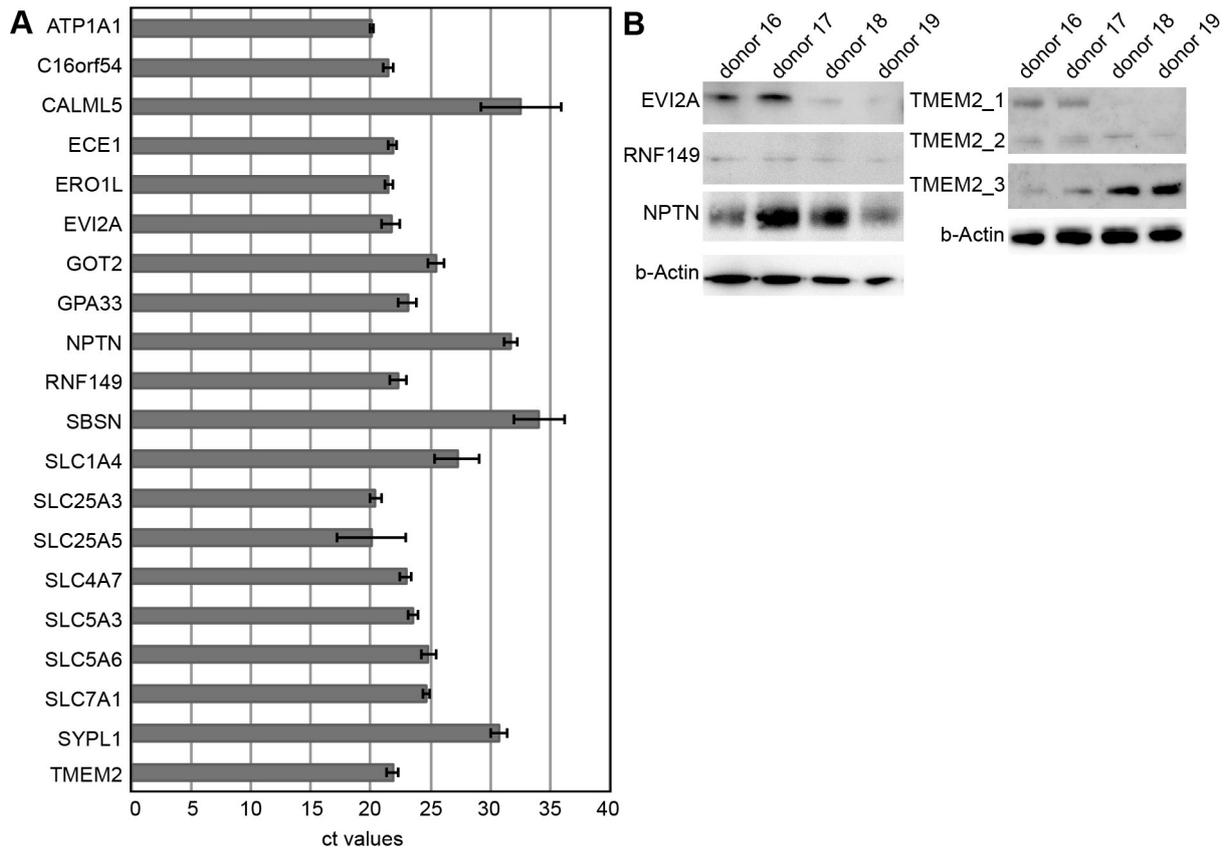
Protein	Description	UniprotID	Peptide counts u.f. quantification	confidence score	UniProt Keywords2GO
APOC3	apolipoprotein C-III	P02656	1	49.94	lipid metabolic process, lipid transport
ATP1A1	sodium/potassium-transporting ATPase subunit alpha-1	P05023	2	58.17	ion transport, nucleotide binding
c16orf54	chromosome 16 open reading frame 54	Q6UWD8	2	53.52	-
CALML5	calmodulin-like protein 5	Q9NZT1	2	61.11	ion binding
DCD	dermcidin	P81605	2	37.72	hydrolase activity
ECE1	endothelin-converting enzyme 1	P42892-4	5	156.03	ion binding, hydrolase activity
ERO1L	ERO1-like protein alpha	Q96HE7	1	21.64	oxidation-reduction process, apoptotic process
EVI2A	ecotropic viral integration site 2A	P22794-2	1	17.62	-
GOT2	glutamic-oxaloacetic transaminase 2	P00505	1	28.63	lipid transport, transferase activity
GPA33	glycoprotein A33	Q99795	2	55.96	receptor activity*
INHBB	inhibin, beta B	P09529	1	23.13	protein binding
NPTN	neuroplastin	Q9Y639	3	58.96	cell adhesion
RNF149	ring finger protein 149	Q8NC42	1	14.53	ligase activity, ion binding
SBSN	suprabasin	Q6UWP8	2	38.45	-
SLC1A4	solute Carrier Family 1, member 4	P43007	1	19.51	alanine, serine, cysteine, and threonine transport
SLC25A3	solute Carrier Family 25, member 3	Q00325	1	18.72	phosphate ion transmembrane transport
SLC25A5	solute Carrier Family 25, member 5	P05141	1	35.3	adenine transport
SLC4A7	solute Carrier Family 4, member 7	Q9Y6M7-7	7	154.69	Sodium bicarbonate cotransport
SLC5A3	solute Carrier Family 5, member 3	P53794	2	39.79	myo-inositol:sodium ion cotransport
SLC5A6	solute Carrier Family 5, member 6	Q9Y289	2	65.01	sodium/multivitamin and iodide cotransporter
SLC7A1	solute Carrier Family 7, member 1	P30825	3	90.99	cationic amino acids (arginine, lysine and ornithine) transporter
SMR3B	submaxillary gland androgen-regulated protein 3B	P02814	1	19.56	-
SYPL1	synaptophysin-like protein 1	Q16563	2	74.95	transport
TMEM2	transmembrane protein 2	Q9UHN6	1	24.65	involved in multicellular organismal development

**Table 3: Summary of the proteins identified by PAL-qLC-MS/MS, which are not co-cited with T cell biology so far.** Protein names and available description of the proteins, which were not co-cited with T cell biology so far. Number of identified peptides, which were used for quantification (pep counts u. f. quantification), and related confidence score resulting from the MS analysis are given, combined with molecular and functional GO terms provided by the UniProt Keywords2GO database. (\* not given as keyword, but PMID:9012807)

At the time the literature analysis was performed, it revealed that 24 (14 %) of the listed proteins were not co-cited with “T cell and/ or activation, proliferation, differentiation” before. A list of these 24 proteins is given in Table 3 with additional information, as far as available, on the description, the UniProt ID and UniProt Keywords2GO. These 24 proteins were also included in the unsupervised cluster analysis (section 1.2.2) of the cell surface glycoproteome dataset (Supplemental Table S4) and Fig. 13 shows selected representatives of these 24 proteins grouped in one of the three clusters according to their expression profiles. Since these proteins were not co-cited with T cell biology before, a validation experiment with two complementing techniques was set up. On the one hand, all membrane-anchored proteins were examined at the transcriptional level by qPCR in the samples of naive CD4<sup>+</sup> T cells of four additional donors (D12-D15) (Fig. 14A). On the other hand, validation on protein level was performed in another four samples of naive CD4<sup>+</sup> T cells from four additional collected donors (D16-D19). Western blots with validated antibodies against four cell surface targets (EV12A, NPTN, RNF149, TMEM2) were performed and demonstrated presence on protein level (Fig. 14B). The expression of the remaining cell surface proteins, which were not co-cited with T cell biology before, could be proven as soon as more validated antibodies are commercially available.



**Figure 13: Dynamic expression profiles of selected cell surface proteins not co-cited with T cell biology so far.** Expression profiles of selected proteins during anti-CD3/anti-CD28 stimulation (0, 3, 6, 12, 24, 48h) (n=4, mean  $\pm$  SEM), grouped according to their allocation in one of the specific clusters of the unsupervised cluster analysis (GProX, see Fig. 11). (modified after <sup>157</sup>)



**Figure 14: Expression validation of selected cell surface proteins not co-cited with T cell biology so far.** Naive CD4<sup>+</sup> T cells were collected from 8 further human blood donors (D12-D19) for validation experiments. A) qPCR analysis of membrane-anchored (n=20) proteins of newly collected naive CD4<sup>+</sup> T cells (n=4, D12-D15). Results are given as ct values of the baseline expression of the protein-corresponding transcripts (mean +/- SEM). B) Western blot analysis of several selected cell surface proteins (EVI2A, RNF149, NPTN, TMEM2) in samples of newly collected naive CD4<sup>+</sup> T cells (n=4, D16-D19). Results are given as images of the Western blots showing the respective bands of each protein and the housekeeper protein beta-Actin (b-Actin) as loading control (TMEM2: There are several existing isoforms, detected by one antibody, and via this Western blot analysis, three isoforms were detected in total with a heterogeneous expression between the biological replicates). (modified after <sup>157</sup>)

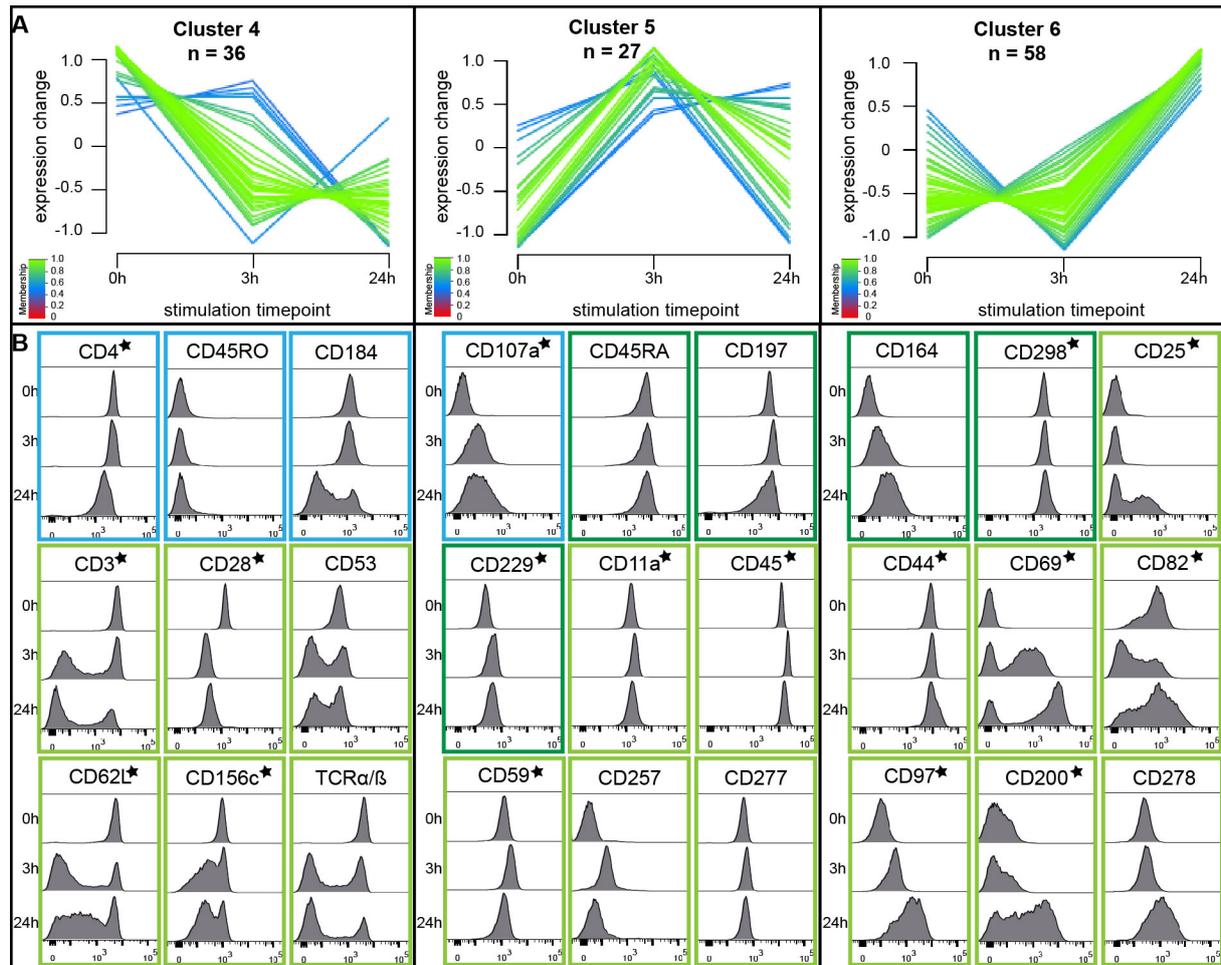
### 1.3 Validation and extension of the cell surface glycoproteome via a targeted flow cytometry-based cell surface screen

To further verify the cell surface glycoproteome dataset, a flow cytometry-based high-throughput screening was carried out. This additionally gave the possibility to identify more cell surface proteins, which were not detected by the PAL-qLC-MS/MS technique.<sup>157</sup> In the course of this, naive CD4<sup>+</sup> T cells were collected from three donors (D5-D7) and stimulated in a time course-dependent manner with anti-CD3/anti-CD28 for 3 and for 24h. These nine samples were subjected to immunostaining with PE-labeled monoclonal antibodies against 332 known cell surface antigens (Fig. 4, orange track). By utilizing a commercially available cell surface screening kit, containing 332 monoclonal antibodies and ten controls, 123 cell

surface markers were detected on the surface of naive and/or activated CD4<sup>+</sup> T cells (Table 2, marked by a filled circle under “identified via”).

### 1.3.1 Cluster analysis of protein expression patterns detected via flow cytometry

To analyze, if the 123 proteins (detected by the flow cytometry screen) showed different expression patterns during the stimulation time course of the naive CD4<sup>+</sup> T cell, an unsupervised expression cluster analysis (Fig. 15A) was performed.<sup>157</sup>

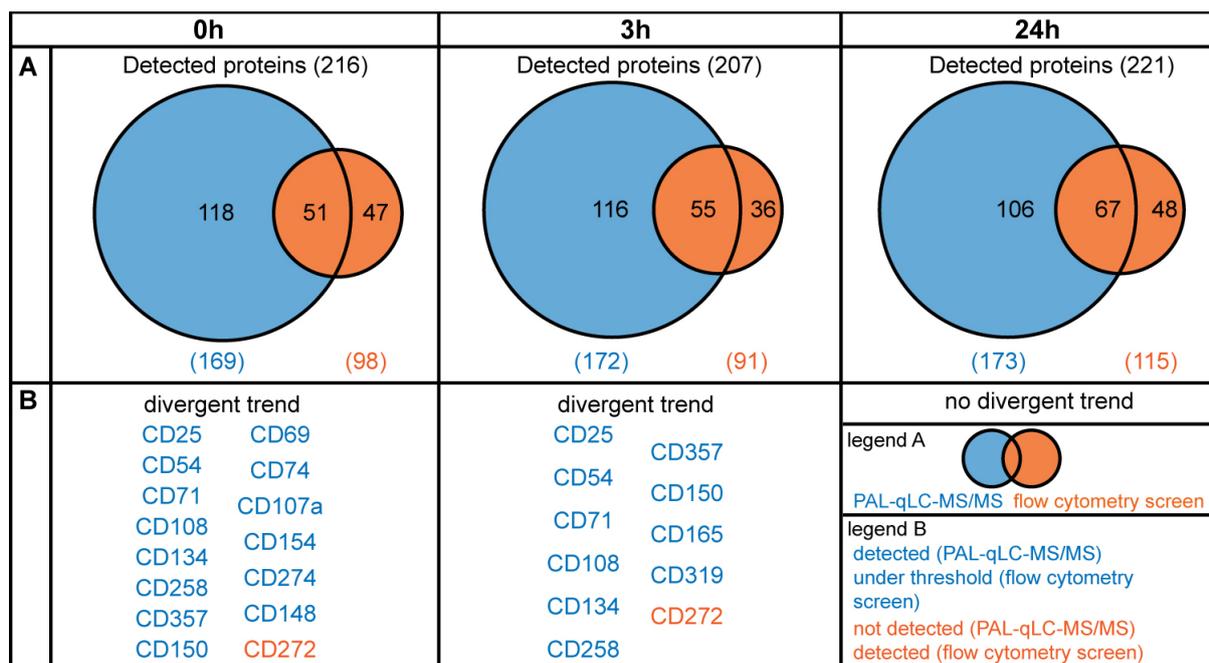


**Figure 15: Unsupervised cluster analysis of the cell surface proteins of human naive and activated CD4<sup>+</sup> T cells, identified by the targeted flow cytometry screening panel.** A) GProX cluster analysis of all cell surface proteins ( $n=123$ ), which were identified via the flow cytometry screening. Clustering is based on the expression profile of the proteins (mean fluorescence intensity, MFI) during the anti-CD3/anti-CD28 stimulation (0, 3, 24h) of one representative donor. The color of each line corresponds to one protein, representing the membership value of this protein, explaining how exact the expression of this protein fits to this respective cluster. Only proteins with a membership value above 0.4 are shown (a table including all membership values for the respective proteins in the different clusters is given in Supplemental Table S5). B) Histograms of the measurements of selected cell surface proteins representing each of the GProX clusters 4-6, detected on naive and/or activated CD4<sup>+</sup> T cells, are shown (results from one representative donor). The color frames of the histograms represent the membership value of the protein to the respective cluster (blue: membership 0.4–0.6; dark green: 0.6–0.8, light green: 0.8–0.99). Cell surface proteins, which are labeled with an asterisk, were also identified via the PAL-qLC-MS/MS approach.<sup>157</sup>

The measured mean fluorescence intensity (MFI) ( $n=2-3$ ) of the detected cell surface proteins was subjected to the unsupervised cluster analysis. This analysis revealed three different clusters (Fig. 15A) according to the detected expression changes during the anti-CD3/anti-CD28 stimulation time course (Supplemental Table S5). This result is in agreement with the result of the cluster analysis of the cell surface glycoproteome (Fig. 11). Fig. 15B depicts histograms of the stainings with the single PE-labeled antibodies of selected representatives for the different clusters, displaying the recorded antibody signals during the T cell activation of one representative donor.

### 1.3.2 Comparison of surface glycoproteome and results of the flow cytometry screen and implementation of flow cytometry screening results into proteomic cell surface atlas

The cell surface glycoproteome, obtained via PAL-qLC-MS/MS, and the results of the flow cytometry screen were compared for the available time points (0, 3 and 24h of stimulation) to investigate similarities and differences.



**Figure 16: Comparison of PAL-qLC-MS/MS and flow cytometry screening panel results.** All cell surface proteins detected by PAL-qLC-MS/MS and/or the flow cytometry screening panel were examined according to their similarity. A) Total number of detected proteins per time point and corresponding technique source are shown in the colored circles, as well as total numbers of detected proteins per technique. B) Proteins detected differently between the two techniques are listed in the color corresponding to the technique source, which identified them in contrast to the other technique, which did not detect the respective protein above the threshold at the time points 0h and 3h. At the 24h time point there were no divergent trends in the detection of the proteins between the techniques. (if a protein was detected in the sample of one donor, this was taken as true for this analysis)

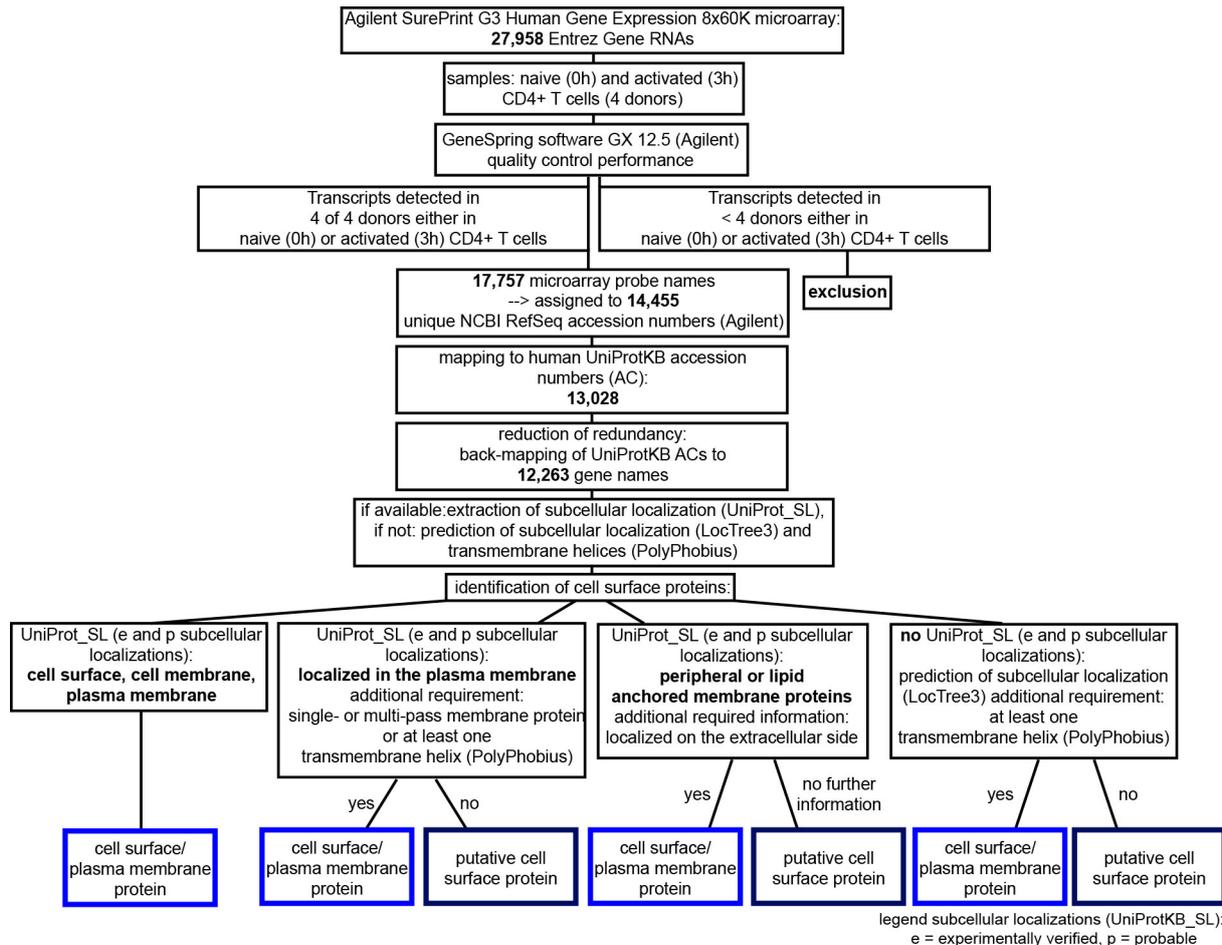
The total number of identified proteins at the different time points is comparable (0h: 216, 3h: 207, 6h: 221). This is also true for the number of detected proteins per technique

(Fig. 16A) and the overlap between the techniques, for the three depicted time points. Divergent trends on the detection of proteins between the techniques could in general be observed in two different ways. One option is that the protein was detected by PAL-qLC-MS/MS and an antibody for this target was included in the cell surface screening kit, but the obtained signal of the antibody staining was under the threshold (Fig. 16B, protein names in blue). The other option is that an antibody gave a signal above the threshold and in general this targeted protein was also detected via PAL-qLC-MS/MS, but not at the same time point as it was detected via the antibody-based analysis (Fig. 16B, protein names in orange). No divergent trend between the two techniques was observed at the 24h time point. The targets, which showed opposing detection trends at the 0 and 3h time points, are nearly the same for both comparisons.

The flow cytometry screening confirmed the expression of 67 cell surface proteins, which were already detected via PAL-qLC-MS/MS. These screening results extended the list of cell surface proteins for naive and activated CD4<sup>+</sup> T cells about 56 proteins, which were uniquely identified via this technique. Combination of the cell surface glycoproteome and the results of the flow cytometry screening led to the cell surface atlas of naive and activated CD4<sup>+</sup> T cells, containing 229 cell surface proteins measured on protein level.<sup>157</sup>

#### **1.4 *In silico* selection and examination of cell surface protein-coding transcripts based on the whole genome microarray analysis**

To further deepen the analysis of cell surface proteins on naive and activated CD4<sup>+</sup> T cells, a broad transcriptomic level was also considered beneath the two proteomic techniques.<sup>157</sup> A whole genome microarray analysis coupled to *in silico* selection of cell surface protein coding transcripts was utilized to investigate the transcriptomic level. Naive CD4<sup>+</sup> T cells and 3h anti-CD3/anti-CD28 stimulated T cells from four donors (D8-D11) were subjected to a whole genome microarray analysis to examine the transcriptomic events during early T cell activation. Fig 17 depicts a decision tree explaining how the microarray dataset was analyzed. In brief, among the detected genes of the genome-wide array, transcripts coding for cell surface proteins were selected by bioinformatics analysis. This analysis started by mapping the detected transcripts to UniProtKB accession numbers and the collection of subcellular localization information (UniProt-SL, experimentally verified or probable UniProtKB annotation) about the mapped proteins, if available. If not available, the subcellular localization of the transcript corresponding protein was predicted by LocTree3 (subcellular localization prediction) and PolyPhobius (transmembrane helix prediction). Following the strict selection criteria of the decision tree (Fig. 17), extended the cell surface atlas of naive and activated CD4<sup>+</sup> T cells about 927 transcripts coding for cell surface proteins (Supplemental Table S3).

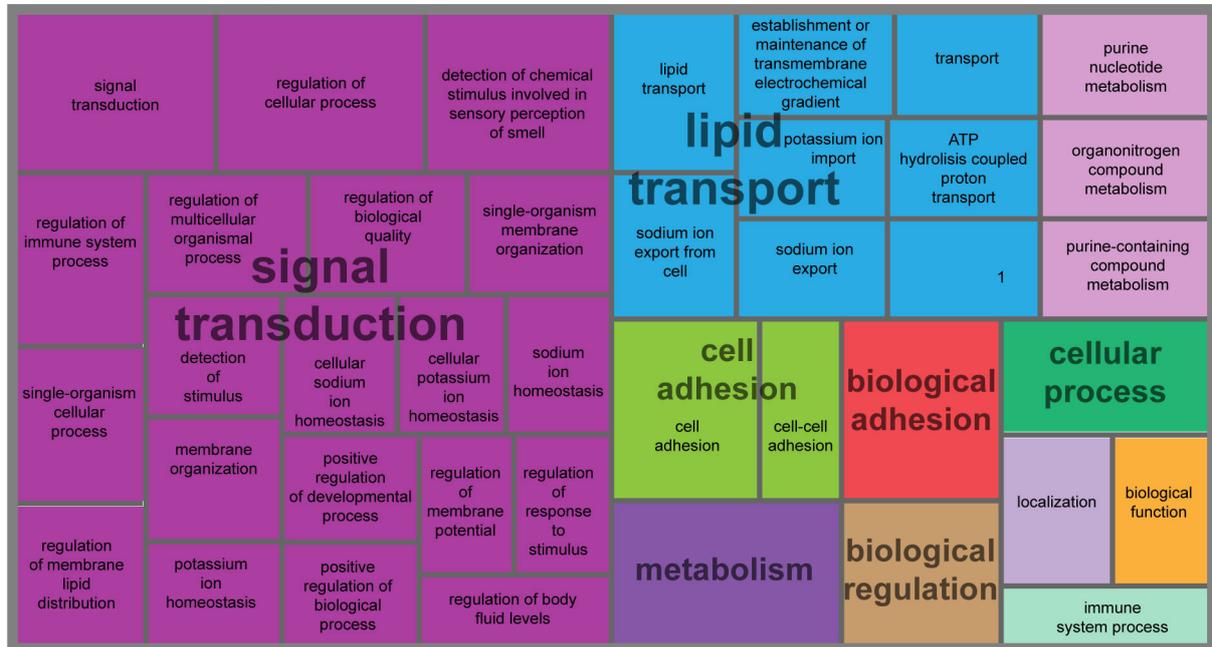


**Figure 17: Decision Tree for the identification of cell surface protein-coding transcripts from the genome-wide transcriptomic microarray analysis.** This decision tree was built to analyze the results of the genome-wide transcription analysis, to be able to *in silico* identify the detected transcripts coding for cell surface- and plasma membrane proteins. In addition it is also possible to identify a set of putative cell surface protein-coding transcripts by following the decision tree.

To get an overview of the entities of the transcriptomic cell surface dataset, a Gene Ontology (GO) enrichment analysis regarding involvement in biological processes as well as molecular function was performed. Enriched GO terms, overlapping between the cell surface protein coding transcripts, were identified with the Generic GO TermFinder algorithm and combined as “superclusters” and visualized as treemaps (REVIGO) (Fig. 18). The GO term analysis of biological processes mainly showed general terms of tasks in which a broad variety of cell surface/plasma membrane proteins could be involved like “signal transduction”, “lipid transport” or “cell adhesion” in contrast to the very specific immune system related GO terms which were seen in the analysis of the PAL-qLC-MS/MS data (Fig. 11B). The GO term analysis of molecular function revealed that one third of the analyzed transcripts shared “nucleoside-binding”-capacity and another third is involved in “primary active transmembrane transporter activity” or in “ATPase activity, coupled”. Both GO term analyses of

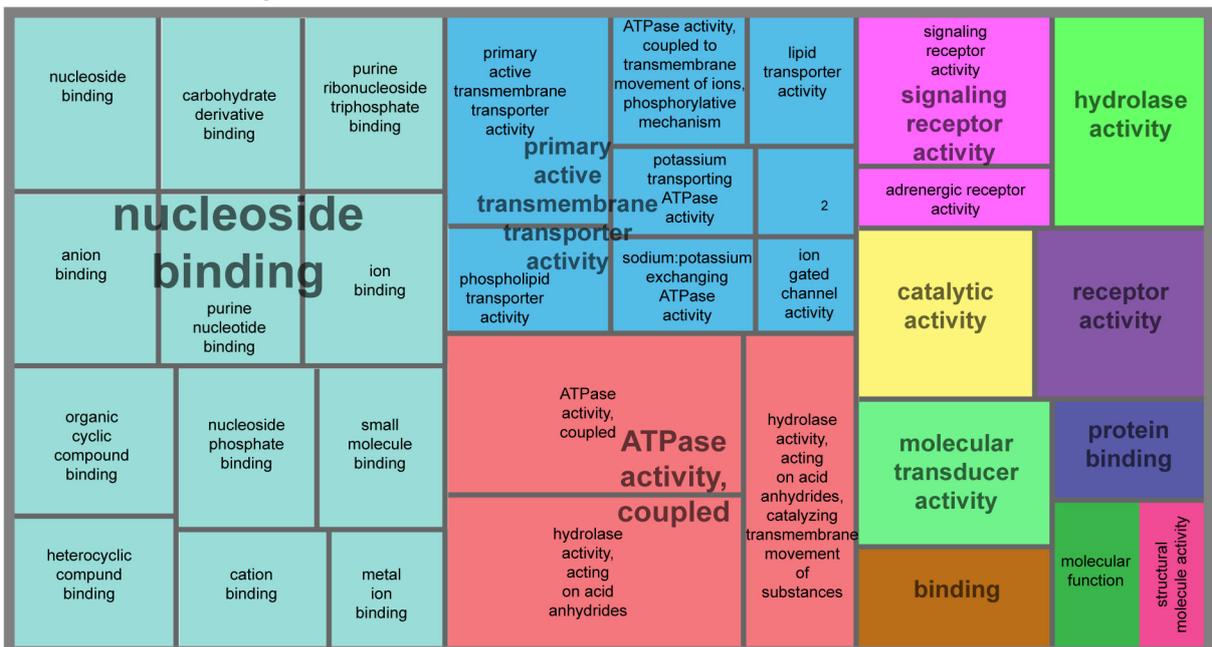
the cell surface protein coding transcripts do not point to an immunological origin of the sample material.

### A. GO-Term Analysis of biological process



1. energy coupled proton transmembrane transport against electrochemical gradient

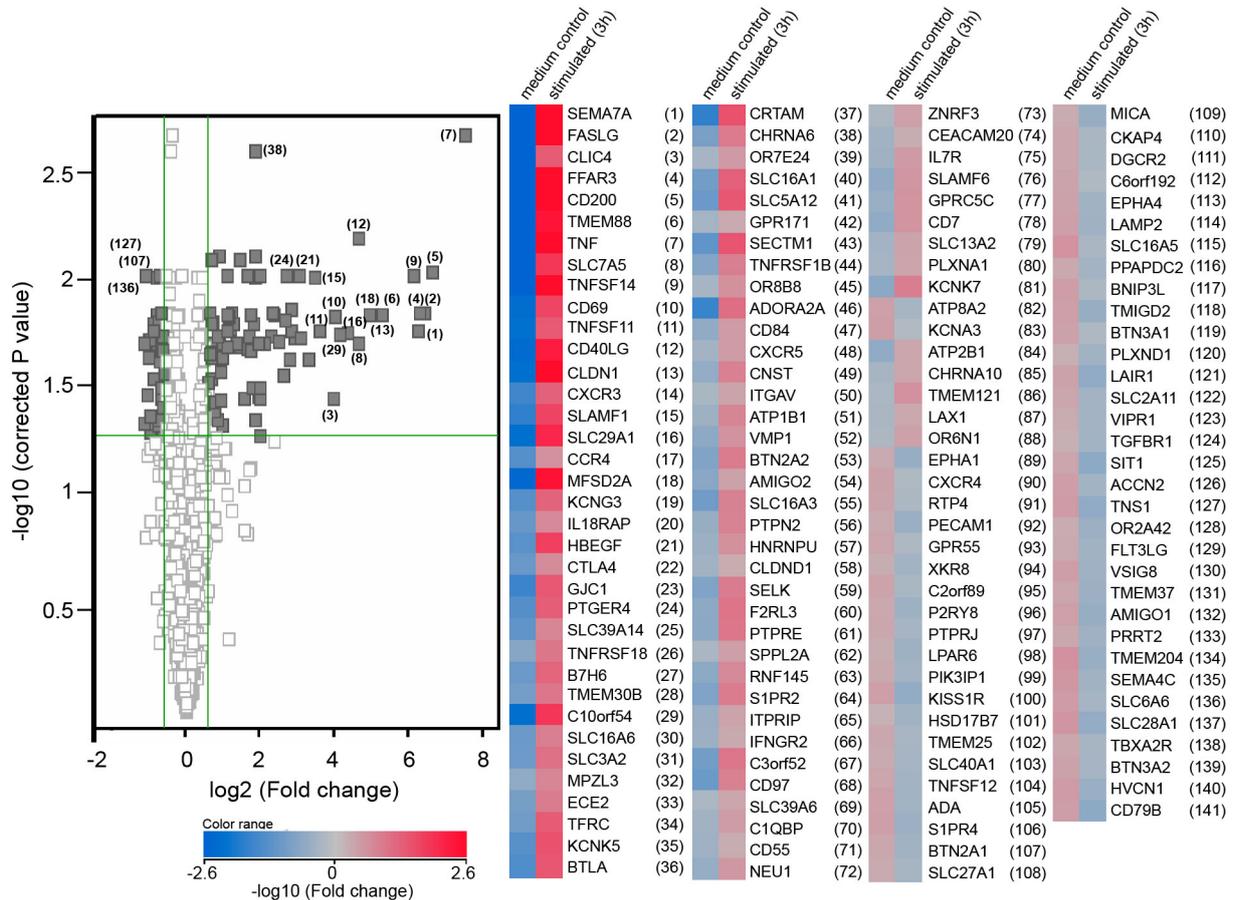
### B. GO-Term Analysis of molecular function



2. intracellular calcium activated chloride channel activity

**Figure 18: GO term enrichment analysis of transcripts coding for cell surface- and plasma membrane proteins from the genome-wide microarray expression analysis.** REVIGO treemaps was used to visualize GO (Gene Ontology) term analysis for the transcripts, which were *in silico* identified to code for cell surface- and plasma membrane proteins (n=927). GO term analyses of biological process (A) as well as molecular function (B) were conducted to get an overview of representative GO terms. Each single rectangle is a representative of an enriched GO term (smaller letters) and these representatives are combined as “superclusters” (larger letters) of loosely connected terms, shown by the different colors. The p-value of the GO term in the underlying GOA database is visualized by the adjusted size of the rectangles.<sup>167</sup>

An analysis of these 927 cell surface coding transcripts, according to differential gene expression between naive and 3h activated CD4<sup>+</sup> T cells, displayed a significant differential expression for 141 genes (Fig. 19). These genes include many targets, which are known to be relevant for immune system- or even especially T cell related processes (SEMA7A, FASLG, CD200, TNF, CD69, CD40LG, CTLA4, BTLA).<sup>157</sup>



**Figure 19: Gene expression analysis of transcripts coding for cell surface- and plasma membrane proteins.** Transcripts coding for cell surface- and plasma membrane proteins ( $n=927$ ), which are expressed in human naive and/or stimulated CD4<sup>+</sup> T cells, were analyzed concerning their differential expression between their state as naive and 3h of anti-CD3/anti-CD28 stimulation. The differential expression of all these transcripts is depicted as a volcano plot (left) indicating the fold change ( $\log_2$ ) against the p-value ( $\log_{10}$ , Benjamini-Hochberg corrected). The vertical lines depict an absolute fold change of 1.5\* (\*if multiple probes were measured per gene, only the highest measured value was included in the analysis) and the horizontal line depicts the corrected significance level of 0.05. 141 of the 927 analyzed transcripts are significantly differential expressed (FC: 1.5, p-value 0.05 (corrected)) between naive and activated (3h) CD4<sup>+</sup> T cells. The heatmap (right) highlights these 141 significantly regulated cell surface/plasma membrane protein-coding transcripts and depicts their corresponding fold changes in the color code from blue (low expressed) to red (high expressed).<sup>157</sup>

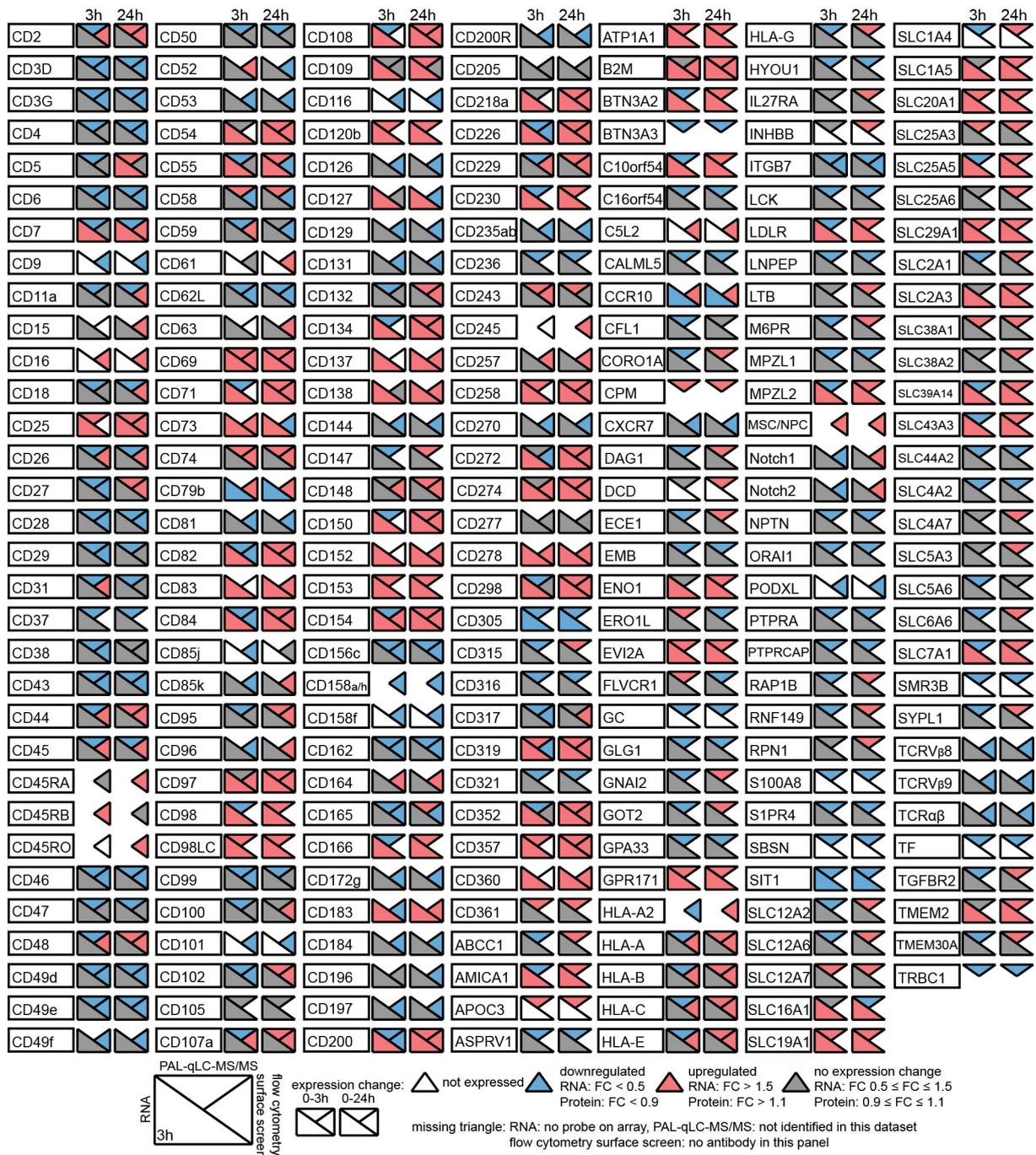
### 1.5 Combination and comparison of proteomic and transcriptomic results

To combine and to analyze the overlap between results seen on proteomic and transcriptomic levels, the three datasets were compared.<sup>157</sup> The investigations on proteomic (PAL-qLC-MS/MS and flow cytometry screening) as well as transcriptomic level (whole ge-

nome microarray analysis coupled to bioinformatics) offered the opportunity to gain a deep insight into the processes happening during naive CD4<sup>+</sup> T cell activation with anti-CD3/anti-CD28. The combination of the three datasets described 229 cell surface antigens, measured on protein level, and 927 transcripts, which encode cell surface proteins. Comparison of the single entities from the proteomic and the transcriptomic results showed, that 53 % of the proteins, identified via PAL-qLC-MS/MS and/or flow cytometry, were also found in the transcriptomic cell surface dataset. To increase the transcriptomic cell surface dataset even more, the strict bioinformatics selection criteria were attenuated. So, in addition a putative transcriptomic cell surface dataset was created. This additional dataset consists of transcripts, which encode proteins described as potentially expressed on the extracellular side of the plasma membrane by the UniProtKB, such as lipid-anchored or peripheral membrane proteins, without any further information on their localization on the extracellular or intracellular side (Supplemental Table S3, putative). 248 further transcripts potentially encoding cell surface proteins, detected in naive and/or activated CD4<sup>+</sup> T cells, are present in this putative dataset. By comparing them with the cell surface proteins identified via one or both of the proteomic techniques, the overlap between transcriptomic and proteomic level could be increased to 58 % (Table 2, marked by a dashed box under “identified via”). If the proteins with a localization annotation (UniProtKB) of only “membrane” (no further information regarding the subcellular localization) would be additionally included, the overlap between transcripts encoding cell surface proteins and cell surface proteins measured on protein level could be increased even more to 83 %.

#### 1.5.1 The surface atlas of human naive and activated CD4<sup>+</sup> T cells

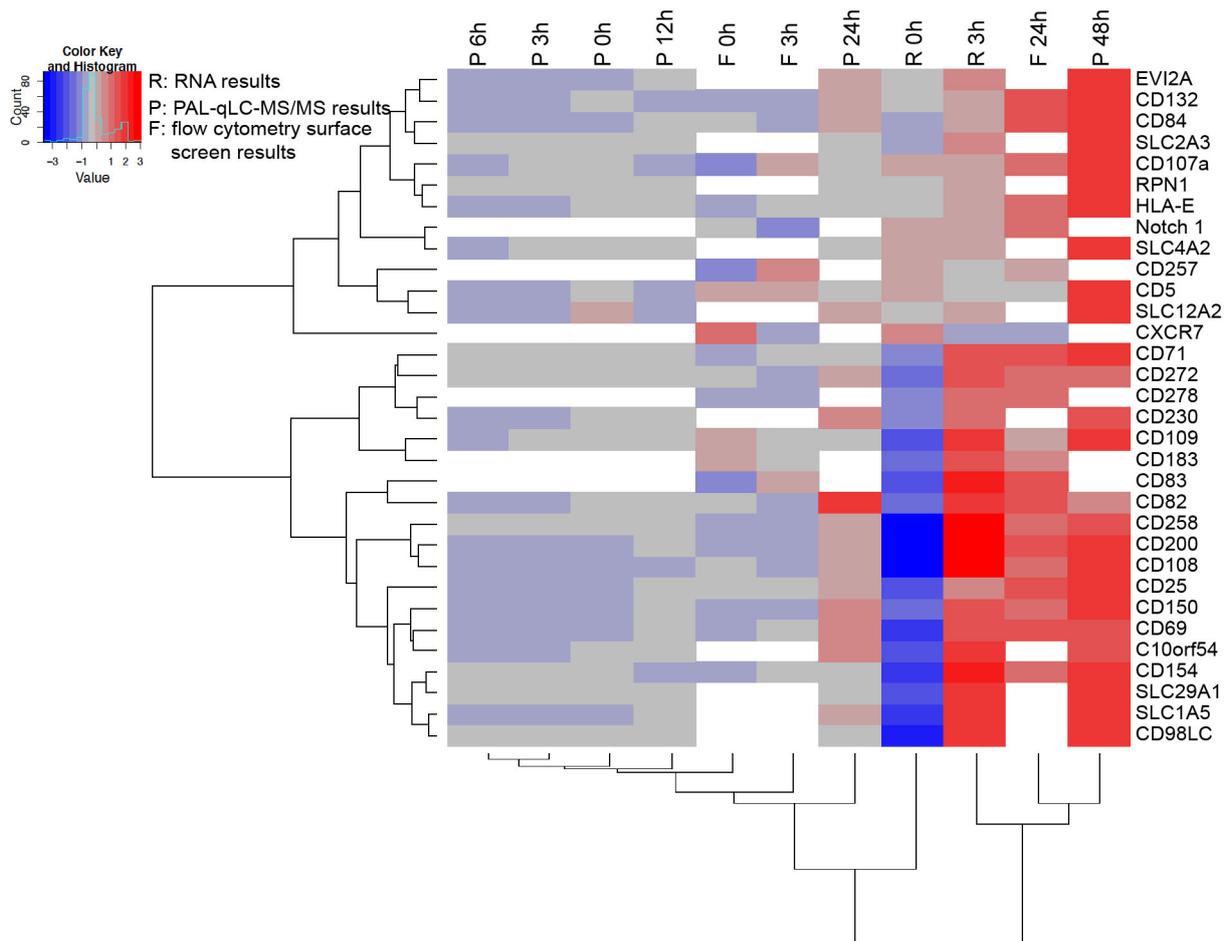
To provide a clear and comprehensive visual presentation of the collected data containing the names of all cell surface proteins and the corresponding expression changes during T cell activation, the cell surface atlas of naive and activated CD4<sup>+</sup> T cells was generated (Fig. 20).<sup>157</sup> Listed here are all the cell surface proteins, which were identified as being present on human naive and/or activated CD4<sup>+</sup> T cells on protein level. Corresponding to each proteins' name, trends of expression change during the activation process measured by PAL-qLC-MS/MS, flow cytometry screen and the transcript expression measured via the genome-wide microarray analysis (mRNA data is independent of prediction or annotation as cell surface protein of the respective transcript) are shown. Fig. 20 presents a reference, aiming to combine all the results of the proteomic and transcriptomic approaches. This atlas drafts the numerous surface proteins and CD markers detected on naive and/or activated CD4<sup>+</sup> T cells, complemented by the direction of their corresponding expression changes during T cell activation.



**Figure 20: The surface atlas of human naive and activated CD4<sup>+</sup> T cells.** All cell surface proteins identified on protein level by the PAL-qLC-MS/MS technique and/or the flow cytometry screen, are depicted in this overview. Also shown are the quantitative trends of protein expression changes after stimulation with anti-CD3/anti-CD28 from 0-3h and 0-24h and their corresponding transcript\* expression changes (\*if multiple probes were measured per gene, only the highest measured value was included in the analysis) from 0-3h.<sup>157</sup>

### 1.5.2 Differential expression analysis of the detected targets from the combined data results of transcriptome and proteome analysis

To be able to select promising surface markers for T cell related functional studies out of the cell surface atlas, the transcriptomic and proteomic results were scaled and combined to find targets significantly regulated on both examined levels.<sup>157</sup> A group of 32 targets, detected and quantified via at least one proteomic and the transcriptomic technique, could be specified (Fig. 21). These 32 targets showed significant regulation in at least one of the techniques for a minimum of one investigated stimulation time point.



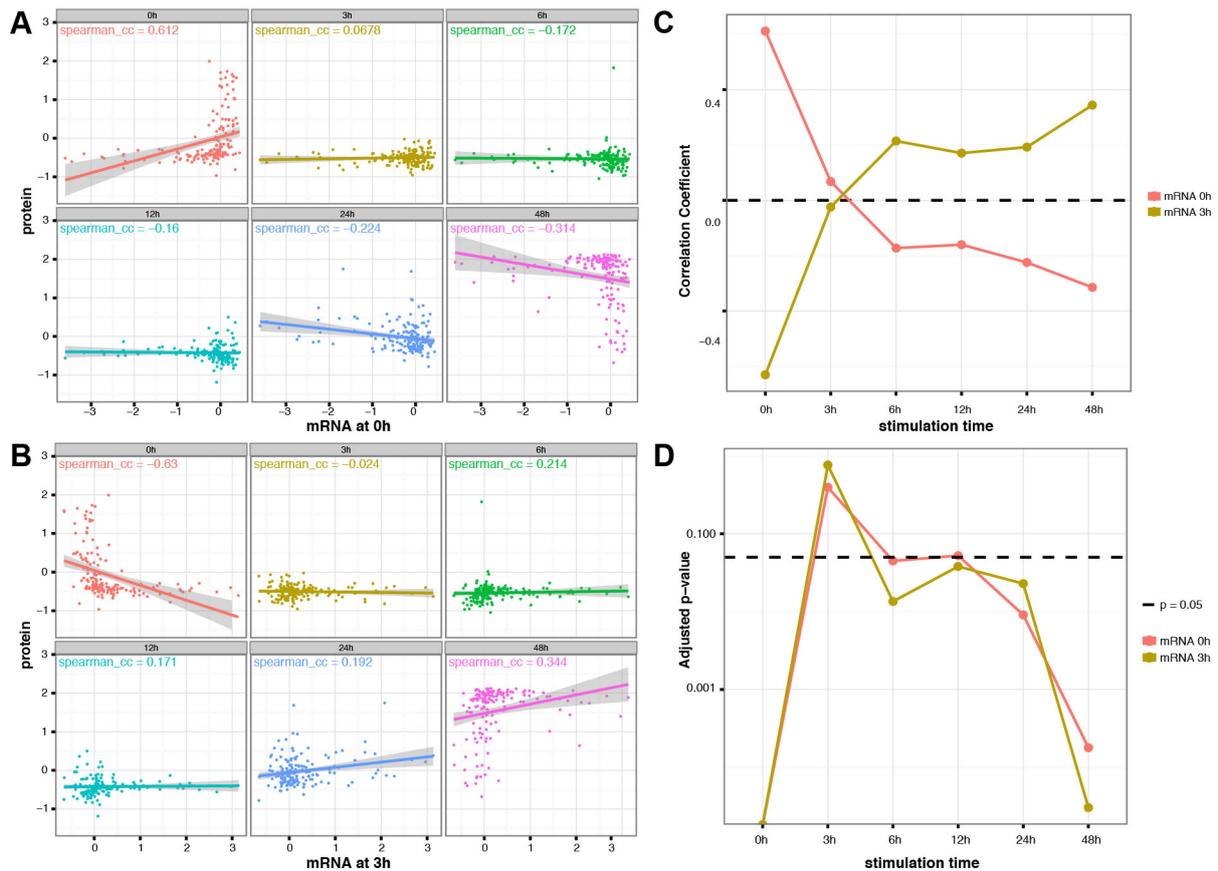
**Figure 21: Heat map of cell surface proteins showing significantly different regulation in the analysis of the combined omics datasets.** Proteins, which were identified on transcriptional as well as on protein level (PAL-qLC-MS/MS and/or flow cytometry screen) were analyzed according to their differential expression ( $FC > 1$ ,  $p < 0.01$ ). 32 cell surface proteins show a significantly differential expression regulation, at least at one of the stimulation time points in at least one of the techniques. Their corresponding expression levels are depicted in a color code scale ranging from blue (low) to red (high).<sup>157</sup>

The largest part of these proteins belongs to the cluster of differentiation (CD), indicating that their role in T cell activation is mostly characterized. Fig. 21 points e.g. also to EVI2A, a cell surface marker, which was not co-cited with T cell biology so far (Table 3). In addition,

the expression of this protein, currently without a known functional role, particularly not T cell related, was confirmed by Western blot (Fig. 14).

### 1.5.3 Correlation of PAL-qLC-MS/MS and whole genome microarray analysis

To deepen the understanding of the relationship between the transcripts and the resulting proteins during early T cell activation, a correlation analysis was performed. By comparing the entities of the proteomic datasets and the strict transcriptomic dataset it was shown that 53 % of the targets are overlapping (section 1.5). Spearman's rank-based coefficient rho was calculated for the targets, which could be measured in all four donors via PAL-qLC-MS/MS as well as on the transcriptomic level. This resulted in the possibility to correlate the expression of 159 targets. Fig. 22A depicts the result of the correlation analysis for the comparison of the transcriptomic target expression at the 0h time point (naive CD4<sup>+</sup> T cell) to the proteomic expression data at 0, 3, 6, 12, 24 and 48h. The highest association could be found for the comparison to the 0h proteomic expression (0.612). The correlation to the following proteomic time points decreases and the second highest association was then shown for the analysis to the 48h proteomic expression, as a weak negative correlation (-0.314). Fig. 22B depicts the correlation calculation of the transcriptomic target expression after 3h anti-CD3/anti-CD28 stimulation (3h activated CD4<sup>+</sup> T cell) to the proteomic expression data at 0, 3, 6, 12, 24 and 48h. The highest Spearman correlation coefficient was calculated for the negative correlation to 0h proteomic expression (-0.63). As for Fig. 20A, the second highest correlation coefficient was calculated for the weak correlation to the 48h proteomic expression (0.334). Fig 22C shows all spearman correlation coefficients from the previous analyses (Fig. 22A and B). This points out that the values of the coefficients were the same for both analyses, but pointing in different directions. The observed correlation coefficients do not provide evidence for strong correlations between transcriptomic and proteomic datasets, only a trend can be seen for both transcriptomic expression time points to the 0h and the 48h proteomic expression time points. The test of the null hypothesis of no correlation led to statistically significant p-values for the association of both transcriptomic expression time points correlated to 0, 6, 12, 24 and 48h proteomic expression.



**Figure 22: Correlation of proteomic (PAL-qLC-MS/MS) to transcriptomic results (microarray expression analysis).** Scaled Quantitative expression values of cell surface proteins, which were detected by PAL-qLC-MS/MS ( $n=4$ , protein abundance) and by the genome-wide microarray expression analysis ( $n=4$ , expression values). A) mRNA values from human naive (0h)  $CD4^+$  T cells were correlated to each of the stimulation time points, which were measured by PAL-qLC-MS/MS (0, 3, 6, 12, 24, 48h) and the corresponding spearman correlation coefficient is given in the respective graph. B) mRNA values from human stimulated (3h)  $CD4^+$  T cells were correlated to each of the stimulation time points, which were measured by PAL-qLC-MS/MS (0, 3, 6, 12, 24, 48h) and the corresponding spearman correlation coefficient is given in the respective graph. C) A summary of the calculated spearman correlation coefficients (A, B) at the corresponding stimulation time points (PAL-qLC-MS/MS). D) The null hypothesis of no association between transcriptomic results (0, 3h) and the proteomic results (0, 3, 6, 12, 24, 48h) was tested and the corresponding p-value was corrected for multiple testing. The graph displays the adjusted p-values for the correlation tests at the different time points and shows that the p-values at the 0, 6, 12, 24 and 48h time points are statistically significant ( $p \leq 0.05$ ).

### 1.6 The distribution of members of the SLC (solute carrier transporters) protein family on naive and activated $CD4^+$ T cells

To deepen the understanding of a potential therapeutic background of the targets, which were identified on the protein level, the proteins were examined about their annotation as drug target (Table 2, DrugBank).<sup>157</sup> The largest group of cell surface proteins within the cell surface atlas is the group of proteins with affiliation to the cluster of differentiation (CD). They represent 55 % of the 229 proteins, most of them described to be present on different cells of the immune system and 29 of them are annotated as targets for approved drugs.<sup>168</sup>

The group with the second highest number of proteins within the atlas is the family of solute carrier (SLC) transporters. In total, 28 members of this family were identified via the PAL-qLC-MS/MS approach and they are listed separately in Table 4 with additional information obtained from UniProt. This protein family is composed of membrane transporters for different kinds of molecules such as ions, sugars and amino acids.

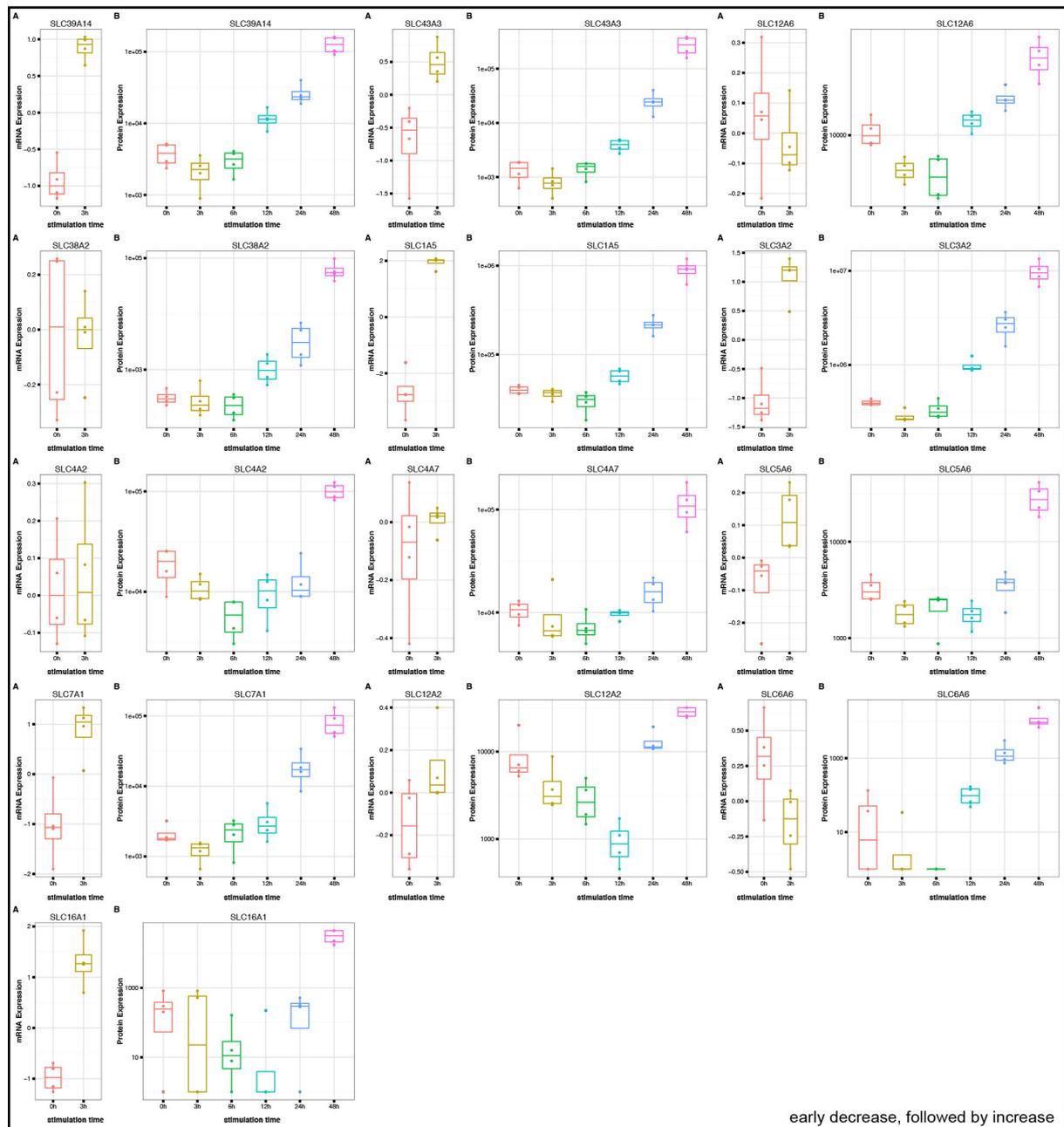
SLC name	UniProt ID	UniProt information	DrugBank ID	Drug name	Drug group	Drug actions
SLC12A2	P55011	basolateral Na-K-Cl symporter, NKCC1	DB00761	Potassium Chloride	1, 2	s
			DB00887	Bumetanide	1	i
			DB01325	Quinethazone	1	i
SLC12A6	Q9UHW9-2	electroneutral potassium-chloride cotransporter 3, KCC3	DB00761	potassium chloride	1, 2	s
SLC12A7	Q9Y666	electroneutral potassium-chloride cotransporter 3, KCC4	DB00761	potassium chloride	1, 2	s
SLC16A1	P53985	MCT1, monocarboxylate transporter 1 (lactate, pyruvate, branched-chain oxo acids derived from leucine, valine, isoleucine, and ketone bodies acetoacetate, beta-hydroxybutyrate and acetate)	DB00175	Pravastatin	1	
			DB00415	Ampicillin	1	
			DB03793	benzoic acid	1	
			DB00119	pyruvic acid	1, 3	
			DB00345	Aminohippurate	1	s
			DB00529	Foscarnet	1	s
			DB00563	Methotrexate	1	s
			DB00936	salicylic acid	1	s
			DB03166	acetic acid	1	s
			DB00313	valproic acid	1, 4	s
			DB00731	Nateglinide	1, 4	s
			DB00627	Niacin	1, 3, 4	s
			DB01032	Probenecid	1	i
			DB04552	Niflumic acid	1	i
DB01440	gamma hydroxybutyric acid	1, 5	i			
DB04398	lactic acid	1	s, i			
DB00119	pyruvic acid	1, 3	s, i			
SLC19A1	P41440	folate transporter 1	DB00563	Methotrexate	1	s
			DB06813	Pralatrexate	1	s
SLC1A4 §	P43007	neutral amino acid transporter a (alanine, serine, cysteine, threonine) sodium dependent	DB00160	L-Alanine	1, 3	
SLC1A5	Q15758	neutral amino acid transporter b (glutamine, asparagine, and branched-chain and aromatic amino acids) sodium-dependent, may also be activated by insulin	DB00130	L-Glutamine	1, 3, 4	s
			DB00174	L-Asparagine	1, 3	
SLC25A5 §	P05141	ADP/ATP translocase 2	DB00720	Clodronate	1, 4	i
SLC25A6	P12236	ADP/ATP translocase 3	DB00720	Clodronate	1, 4	i
SLC29A1	Q99808	equilibrative nucleoside transporter 1 (influx and efflux of nucleosides), sodium dependent	DB00898	Ethanol	1	
			DB00900	Didanosine	1	
			DB00943	Zalcitabine	1	
			DB00441	Gemcitabine	1	s
			DB00544	Fluorouracil	1	s
			DB00811	Ribavirin	1	s
			DB00987	Cytarabine	1, 4	s
			DB01033	Mercaptopurine	1	s
			DB01073	Fludarabine	1	s
DB00642	Pemetrexed	1, 4	in			
SLC2A1	P11166	GLUT-1, facilitated glucose transporter member 1, constitutive or basal glucose uptake, can transport a wide range of aldoses (pentoses and hexoses)	DB00292	Etomidate	1	i
SLC44A2	Q8IWA5	choline-transporter-like protein2	DB00122	Choline	1, 3	s
SLC4A7 §	Q9Y6M7-7	sodium bicarbonate cotransporter 3 (electroneutral sodium-bicarbonate-dependent cotransporter Na <sup>+</sup> : HCO <sub>3</sub> <sup>-</sup> 1:1)	DB01390	sodium bicarbonate	1	s

SLC name	UniProt ID	UniProt information	DrugBank ID	Drug name	Drug group	Drug actions
SLC5A6 §	Q9Y289	sodium-dependent multivitamin transporter, transports pantothenate, biotin and lipoate in the presence of sodium	DB00121	Biotin	1, 3	
			DB00166	Lipoic acid	1, 3	
			DB08872	gabapentin enacarbil	1	s
SLC7A1 §	P30825	high affinity cationic amino acid transporter 1 (arginine, lysine, ornithine)	DB00123	L-Lysine	1, 3	
			DB00125	L-Arginine	1, 3	
			DB00129	L-Ornithine	1, 3	
SLC7A5	Q01650	CD98LC, LAT-1, large neutral amino acids transporter small subunit 1, sodium-independent, high affinity transport of large neutral amino acids (phenylalanine, tyrosine, leucine, arginine, tryptophan when associated with SLC3A2	DB00297	Liothyronine	1	
			DB00451	Levothyroxine	1	
			DB00509	Dextrothyroxine	1	
			DB01042	Melphalan	1	
			DB01235	Levodopa	1	
SLC38A5	Q8WUX1	CD165, sodium-coupled neutral amino acid transporter 5 (glycine, asparagine, alanine, serine, glutamine and histidine)				
SLC20A1	Q8WUM9	sodium-dependent phosphate transporter 1, sodium-phosphate symporter				
SLC25A3 §	Q00325	phosphate carrier protein				
SLC2A3	P11169	GLUT-3, facilitated glucose transporter member 3, can also mediate uptake of other monosaccharides (not fructose)				
SLC38A1	Q9H2H9	sodium-coupled neutral amino acid transporter 1, cotransport of glutamine and sodium ions 1:1				
SLC38A2	Q96QD8	sodium-coupled neutral amino acid transporter 2				
SLC39A14	Q15043-2	zinc transporter zip14, may mediate cellular uptake of nontransferrin-bound iron				
SLC43A3	Q8NBI5	purine-selective nucleobase transporter, highly expressed in macrophages				
SLC4A2	P04920-3	anion exchange protein2				
SLC5A3 §	P53794	sodium/myo-inositol cotransporter, prevents high intracellular accumulation of myo-inositol				
SLC6A6	P31641	sodium-and-chloride-dependent taurine transporter (taurine and beta-alanine transporter), chloride ions are necessary for optimal uptake				
SLC3A2	P08195-2	CD98, 4F2 cell-surface antigen heavy chain, required for function of light-chain amino acid transporters, high affinity transport of large neutral amino acids				

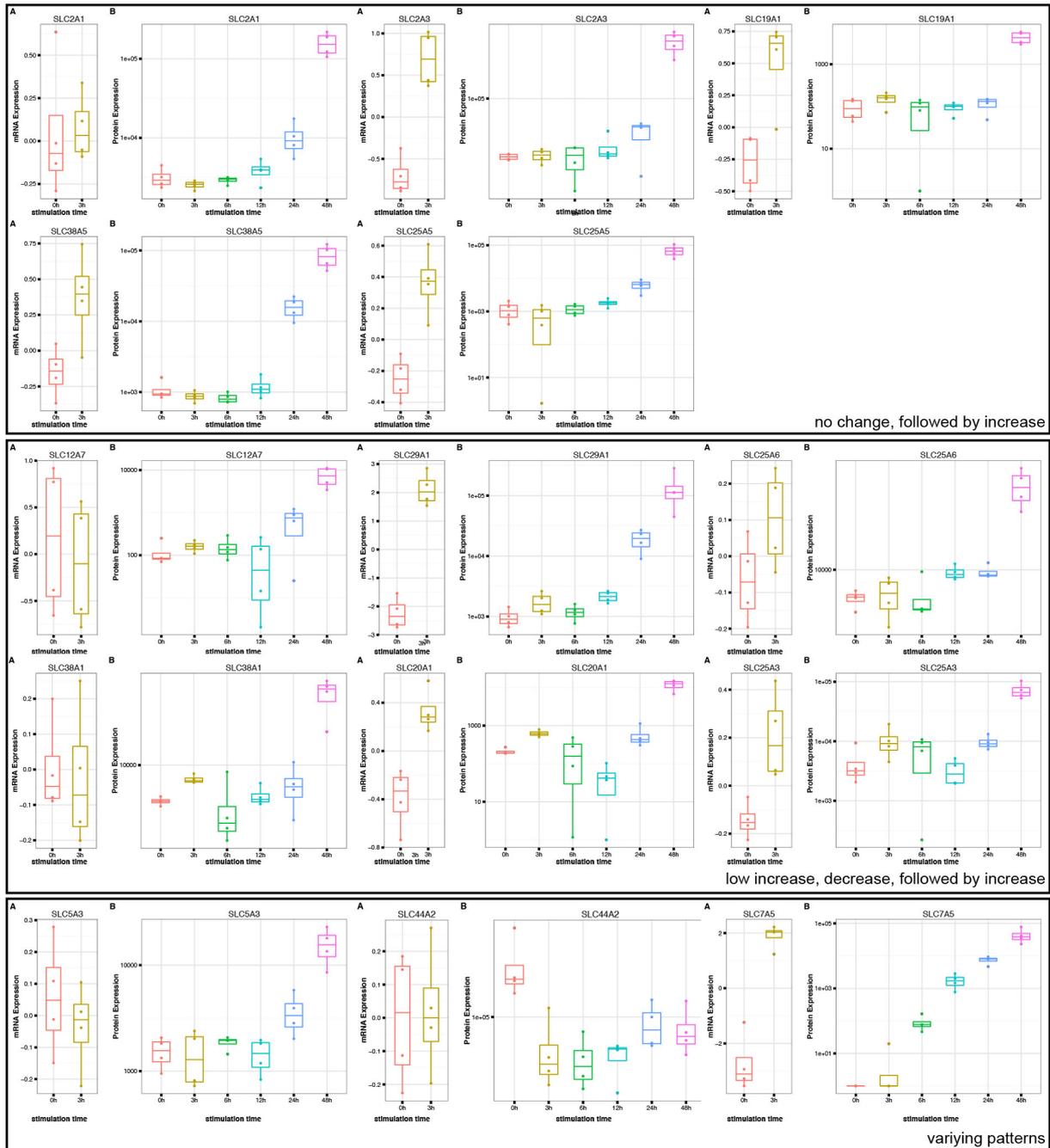
**Table 4: Solute carrier (SLCs) transporters identified via PAL-qLC-MS/MS.** This table presents a list of all SLCs that were included in the cell surface atlas of naive and activated CD4<sup>+</sup> T cells with additional information obtained from UniProt and DrugBank. (§=not co-cited with T cell biology so far; Drug group: 1) approved, 2) withdrawn, 3) nutraceutical 4) investigational, 5) illicit; Drug action: i=inhibitor, s=substrate, in=inducer)

These transporters are diversely expressed on naive CD4<sup>+</sup> T cells. They also exhibit very distinct expression patterns during the stimulation with anti-CD3/anti-CD28. Fig. 23 and 24 show all SLCs, which were detected on mRNA as well as on protein level (n=27, because SLC1A4 was detected in less than four donors on mRNA level). It is possible to compare their relative protein abundance to each other because of the quantitative results of the mass spectrometry-based measurements. Most of the SLCs have a low expression on the naive CD4<sup>+</sup> T cell. SLC4A2, SLC1A5, SLC3A2 (Fig. 23) and SLC44A2 (Fig. 24) are the highest expressed among the depicted SLCs. All of the SLCs (Fig. 23 and 24), besides SLC44A2, exhibit their highest expression after 48h stimulation, suggesting them as interesting targets on activated CD4<sup>+</sup> T cells. On mRNA level there are three different groups: induced upon stimulation (n=19), no change upon stimulation (n=4) and decrease upon stimulation (n=4). The comparison of mRNA and protein level again indicated that the expression changes upon stimulation do not show the same results for all targets (Fig. 23 and 24, each A and B), as also shown by the correlation of the PAL-qLC-MS/MS results to the corresponding RNA results (Fig. 22). But still, these proteins are very interesting because many members of this family are known targets for supplements and drugs. A DrugBank<sup>168</sup> target search revealed

that within the 28 identified SLCs, 16 transporters are registered as drug targets, described in the following categories: approved, investigational, nutraceutical (some of them were withdrawn or are illicit, although approved before) (Table 4).



**Figure 23: Expression patterns (1) of solute carrier (SLC) transporters obtained via genome-wide microarray analysis (A) and PAL-qLC-MS/MS (B).** This figure presents the SLCs, which were identified on mRNA (n=4) as well as on the protein level (n=4), exhibiting an early decrease in protein expression followed by an increase during the stimulation time course. Shown are boxplots with median and points for the single donor values. To be able to compare the pattern on mRNA as well as on protein level, the corresponding results are plotted next to each other.



**Figure 24: Expression patterns (2) of solute carrier (SLC) transporters obtained via genome-wide microarray analysis (A) and PAL-qLC-MS/MS (B).** This figure presents 3 different groups of SLCs, which were identified on mRNA ( $n=4$ ) as well as on the protein level ( $n=4$ ). Shown are boxplots with median and points for the single donor values. To be able to compare the pattern on mRNA as well as on protein level, the corresponding results are plotted next to each other. The first box shows proteins, which do not exhibit any changes in abundance during the first hours of anti-CD3/anti-CD28 stimulation, but increase in expression between 12 and 24h. The second box illustrates SLCs characterized by an expression increase after 3h of stimulation directly followed by a decrease until 12-24h. Most of the SLCs exhibit the highest abundance at 48h of stimulation. The third box shows the three remaining SLCs, which could not be added to any of the previous expression groups.

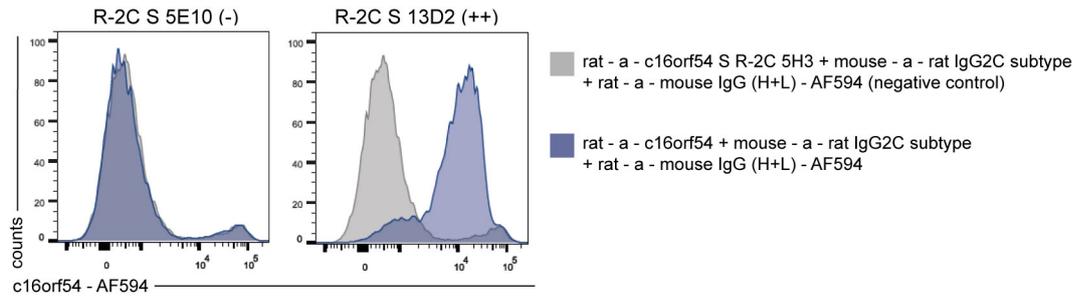
## **2. c16orf54 – a cell surface protein which was not described in the context of T cell biology before**

As mentioned in section 1.2.4, the non-targeted PAL-qLC-MS/MS approach led to the identification of 24 cell surface proteins, which were not co-cited with T cell biology at the time of data collection. For some of them there is no knowledge about the general local distribution within the human system and also no reference for their potential function, so far.<sup>157</sup> One of these candidates is the transmembrane protein c16orf54 (abbreviated as c16orf54), which is encoded by the chromosomal open reading frame 54 on chromosome 16. The predicted molecular weight of the protein is approximately 25 kDa (UniProtKB). This protein was demonstrated to be present on human naive CD4<sup>+</sup> T cells and is characterized by a fast expression decrease upon stimulation with anti-CD3/anti-CD28. After 6h of stimulation the protein abundance of c16orf54 was very low compared to 0h and stayed at this level until 24h of stimulation. After this time point the protein abundance increased again and reached a level similar to that on naive CD4<sup>+</sup> T cells. On transcriptomic level, this expression pattern was different. In naive CD4<sup>+</sup> T cells the transcript of c16orf54 is present and upon stimulation (3, 6, 12, 24, 48h) with anti-CD3/anti-CD28 it constantly decreases (not shown here).

To be able to further investigate the transmembrane protein c16orf54 in the human and murine system, monoclonal antibodies against peptides of the extracellular domain of the protein were generated.

### **2.1 Suitability screening of anti-mouse- and anti-human c16orf54 antibodies (hybridoma supernatants) for flow cytometry**

The Institute of Molecular Immunology of the Helmholtz Center Munich generated the hybridoma supernatants. The supernatants were tested in an ELISA for the presence of antibodies, which are specific for the human and/or murine peptides, which were used to immunize the rats and mice. In a second ELISA the antibody isotype/isotypes in the functional hybridoma supernatants were determined, trying to exclude the possibility that one hybridoma supernatant contains more than one antibody. Afterwards, the hybridoma supernatants were tested in flow cytometry stainings, using a 3-step staining protocol. Murine (Fig. 25) as well as human (Fig. 26) naive CD4<sup>+</sup> T cells were stained using the hybridoma supernatants listed in Table 5 (murine) and Table 6 (human).



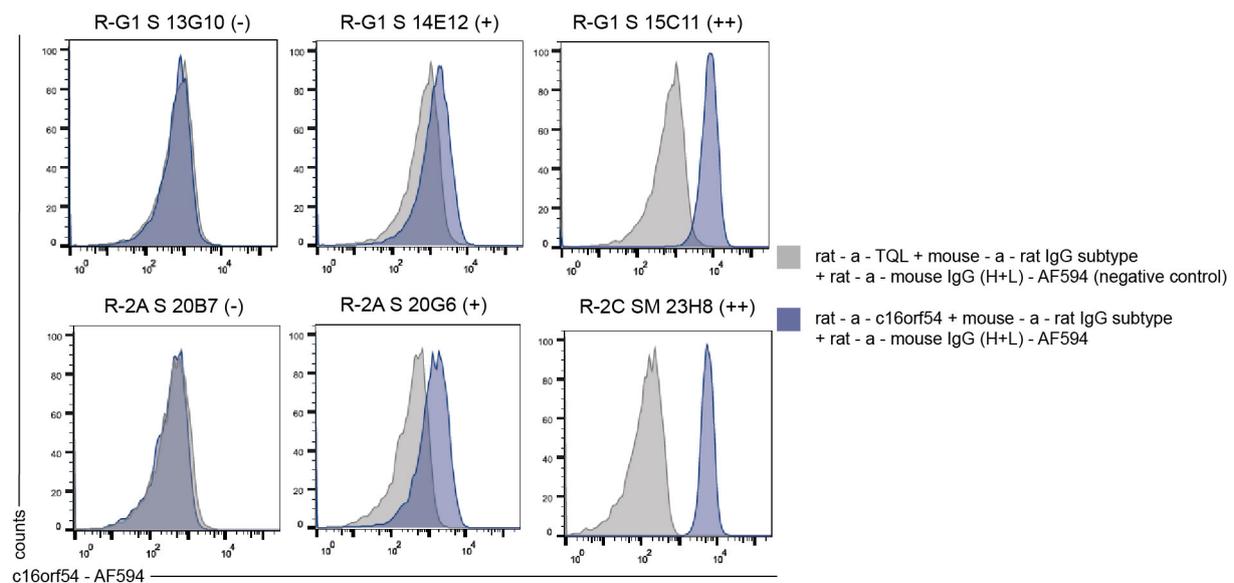
**Figure 25: Flow cytometry suitability screen of hybridoma supernatants (containing monoclonal antibodies against c16orf54) for their specificity for murine naive CD4<sup>+</sup> T cells.** To test the generated hybridoma supernatants, a flow cytometry staining of isolated murine naive CD4<sup>+</sup> T cells was performed (87 hybridoma supernatants containing antibodies of different rat and mouse IgG subtypes). Depicted here are two exemplary histograms showing the results of the flow cytometry staining using the supernatant of hybridoma clone R-2C S 5E10, which did not show a positive staining, and R-2C S 13D2 resulting in a positive signal by using them as primary antibody for the staining procedure. The hybridoma supernatant from clone R-2C 5H3 was used as negative control, because this supernatant only reacted with the human peptide and not the murine peptide in the ELISA.

clone	rating	clone	rating	clone	rating
7F11	-	11B6	-	5D6	-
5H3	-	15F10	-	4H2	-
24F5	-	22G5	-	10A9	-
13D11	-	6A8	-	16A2	-
14H5	-	6B5	-	23A1	-
20G4	-	6H7	-	13D2	++
17F4	-	7H5	-	4A2	-
23D5	-	5H3	-	15E9	-
24G4	-	3H2	-	16C10	-
8B4	-	13B1	-	6D6	-
14H11	-	5A6	-	4E10	-
4D11	-	6A6	-	23H6	-
12F9	-	15D10	-	16G6	-
22G10	-	24F12	-	24F12	-
7F9	-	6B3	-	15C2	-
6B3	-	1A11	-	16D2	-
13A9	-	8A1	-	3G10	-
19B7	-	14F9	-	6A6	-
2C7	-	5G9	-	31H10	-
12A1	-	3B10	-	28G4	-
16A9	-	24C1	-	30D2	-
7C9	-	1H10	-	27G3	-
1G7	-	13D10	-	30C2	-
23H8	-	5E10	-		
1E12	-	24E1	-	<b>Color code:</b>	
15C12	-	22C4	-	rat IgG2A	
16F3	-	14C9	-	rat IgG2C	
15D3	-	22H1	-	rat IgG1	
1B12	-	17A11	-	mouse IgG1	
2F5	-	8C1	-	mouse IgG2a	
6D5	-	24F10	-	mouse IgG2b	

**Table 5: Rating of hybridoma supernatant performance in flow cytometry staining (test on murine cells).** The supernatants of the listed hybridoma clones were tested for their suitability as primary antibodies in flow cytometry stainings. Different antibody isotypes are labeled by a color code. Rating is from “no positive signal” (-) to “precisely distinguishable positive signal” (++)

The results of the staining of murine naive CD4<sup>+</sup> T cells using the hybridoma supernatants showed, that only the hybridoma clone R-2C 13D2 produced an antibody suitable for the use in flow cytometry stainings. Afterwards, the Institute of Molecular Immunology started the subcloning procedure to guarantee the production of a monoclonal antibody. The subcloned 13D2 antibody exhibited similar results in the flow cytometry staining (not shown) and was therefore purified for further testing.

The test of the supernatants in the flow cytometry staining of human naive CD4<sup>+</sup> T cells showed a slightly different result (Fig. 26). In this case it was possible to distinguish three types of staining intensities, ranging from “no signal” (-) to “precisely distinguishable positive signal” (++) . The rating of all tested hybridoma clones is listed in Table 6. The results for the (++) clones were confirmed in a second experiment and after this the clones 7F11 and 23H8 were chosen for subcloning and the purification procedure. The subcloned and purified 23H8 (=23H8\*) antibody showed a better performance than 7F11\* and therefore 23H8\* was chosen for direct labeling with the fluorophore PE-TexasRed.



**Figure 26: Flow cytometry suitability screen of hybridoma supernatants (containing monoclonal antibodies against c16orf54) for their specificity for human naive CD4<sup>+</sup> T cells.** To test the generated hybridoma supernatants, a flow cytometry staining of isolated human naive CD4<sup>+</sup> T cells (110 hybridoma supernatants containing antibodies of different rat IgG subtypes) was performed. Depicted here are histograms showing the results of the flow cytometry stainings from six exemplary hybridoma clone supernatants, which were used as primary antibodies for the staining procedure. The rating ranges from: “no positive signal” (-), “overlapping positive signal” (+), until “precisely distinguishable positive signal” (++) .

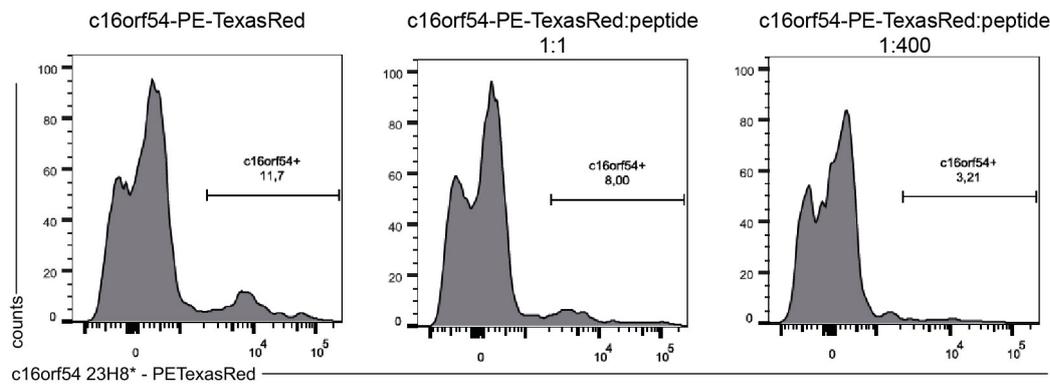
clone	rating	clone	rating	clone	rating
12F9	-	7C6	-	7A8	+
13B1	-	8G11	-	7F10	+
13D11	-	14E2	-	15C12	++
14H11	-	17B6	-	1G7	++
14H5	-	20A8	-	23H8	++
15D10	-	4G11	-	7F11	++
17F4	-	15H9	+	10F7	++
20G4	-	20G6	+	12E4	++
23D5	-	1E12	-	15F7	++
24F5	-	22G5	-	17H9	++
24G4	-	11B6	+	18E13	++
4D11	-	15D3	+	1B1	++
5A6	-	15F10	+	21B1	++
5H3	-	1B12	+	24G3	++
6A6	-	2F5	+	6B4	++
7F11	-	3H2	+	13E12	-
8B4	-	5H3	+	13G10	-
13F11	-	6A8	+	14H10	-
14H6	-	6B5	+	14H6	-
15B3	-	6D5	+	15B9	-
15E4	-	6H7	+	22C9	-
16E9	-	7H5	+	23G5	-
18B7	-	12F6	+	2H6	-
18F6	-	14G6	+	4A6	-
18H8	-	16A3	+	6C12	-
1A4	-	16F4	+	9A10	-
20B7	-	16F8	+	9G8	-
23B11	-	17E10	+	9G9	-
23B3	-	21C4	+	14E12	+
24E11	-	22C5	+	18H8	+
24F2	-	22E1	+	22H5	+
2D6	-	23B11	+	5B4	+
3E3	-	23G10	+	6C2	+
3F2	-	23H7	+	15C11	++
3F4	-	24G9	+	<b>Color code:</b>	
4E2	-	2H5	+	rat IgG2A	
5B10	-	2H6	+	rat IgG2C	
6F7	-	4B1	++	rat IgG1	

**Table 6: Rating of hybridoma supernatant performance in flow cytometry staining (test on human cells).** The supernatants of the listed hybridoma clones were tested for their suitability as primary antibodies in flow cytometry stainings. Different antibody isotypes are labeled by a color code. Rating is from “no positive signal” (-), “overlapping positive signal” (+), until “precisely distinguishable positive signal” (++)

## 2.2 Assessment of anti-human c16orf54 antibody specificity and distribution of c16orf54 on human PBMCs

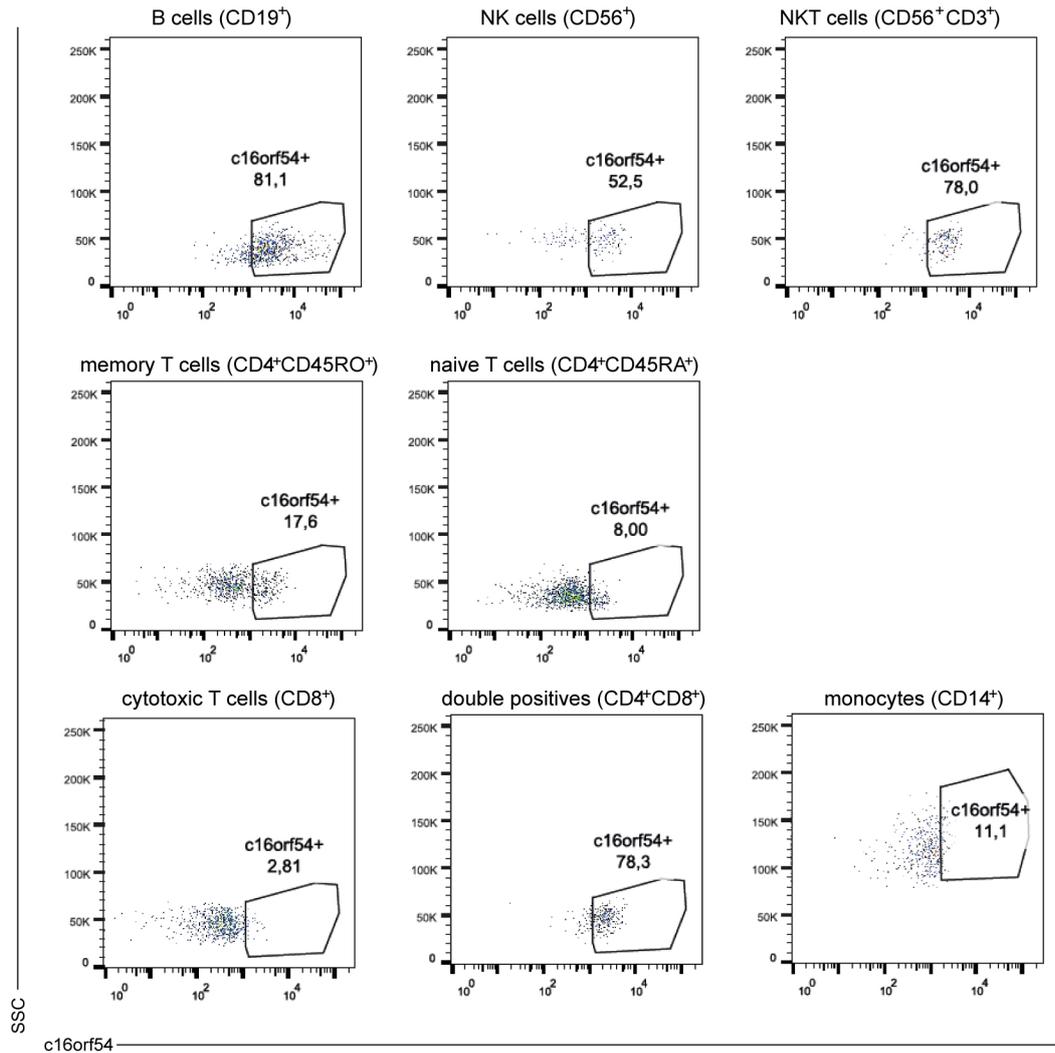
To assess the specificity of anti-human c16orf54 23H8\*-PE-TexasRed, a competition assay was carried out. Anti-c16orf54 23H8\*-PE-TexasRed was used alone or pre-incubated with varying amounts of the peptide, which was used for immunization, for the staining of human

PBMCs. Fig. 27 depicts the staining signal, which was measured by flow cytometry. A slight signal decrease is seen between the staining using the antibody only (Fig. 27, left panel) and an antibody to peptide ratio of 1:1 (middle panel). A 400 x excess of peptide nearly diminished the signal of the antibody, confirming its specificity for the extracellular part of c16orf54.



**Figure 27: Antibody-peptide competition staining.** To test the specificity of the purified and labeled antibody clone c16orf54 23H8\*-PE-TexasRed a competition assay coupled to flow cytometry analysis was performed. Different ratios of antibody:peptide were pre-incubated for 30min at RT and this mixture was used as staining solution for human PBMCs and analyzed via flow cytometry.

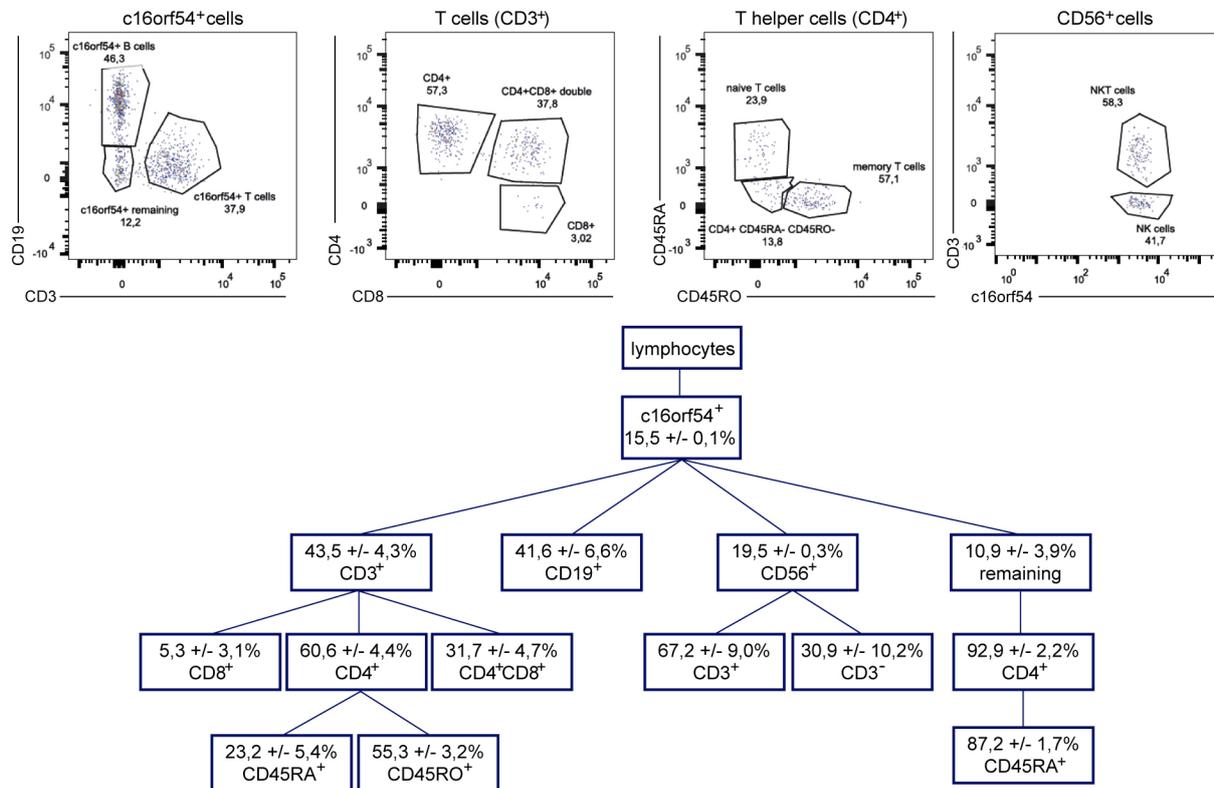
Many proteins are not exclusively expressed on one type of cell, therefore the distribution of c16orf54 expression on human PBMCs was checked via flow cytometry. Fig. 28 shows the results of the flow cytometry analysis of human PBMCs, which were stained with known cell type specific antibodies and c16orf54 23H8\*-PE-TexasRed. The plots display the gated cell populations as indicated in each heading (one representative donor). Shown is the gate for c16orf54<sup>+</sup> cells, complemented with the percentages of positive cells within the mother population. Based on this it seems that only a small percentage of human naive CD4<sup>+</sup> T cells is positive for c16orf54 (8 %) and besides this, nearly all B cells (81 %), a fraction of NK cells (52.5 %) and NKT cells (78 %). Only a small fraction of memory T cells (17.8 %) and cytotoxic T cells (2.8 %) also expresses c16orf54.



**Figure 28: Flow cytometry staining of human PBMCs to examine c16orf54 expression on lymphocyte subpopulations.** Human PBMCs were stained with specific antibodies against common lymphocyte cell surface markers (CD3, CD19, CD56, CD4, CD8, CD45RA, CD45RO, CD14) and the purified and labeled antibody c16orf54 23H8\*-PE-TexasRed. The plots show the expression levels of c16orf54 on the previously gated cell populations. (n=3, plots are shown for one representative donor)

To analyze the distribution of the different cell types within the c16orf54<sup>+</sup> population on a percentage basis, a different gating strategy was applied, which is illustrated in Fig. 29. In this analysis the first gate was applied to the c16orf54<sup>+</sup> cell population and starting from there, the PBMC subpopulations were gated. The resulting cell population percentages were calculated for three donors and the lower panel of Fig. 29 depicts how the cell subtypes can be divided from each other and shows their distribution within the c16orf54<sup>+</sup> cell population. The major populations carrying c16orf54 were, T cells (CD3<sup>+</sup>) and B cells (CD19<sup>+</sup>) (both around 40 %), NK cells (CD56<sup>+</sup>/CD3<sup>-</sup>) (6 %) and NKT cells (CD56<sup>+</sup>/CD3<sup>+</sup>) (13 %). The included staining antibodies also gave the possibility to differentiate T cell subsets within the c16orf54<sup>+</sup> population, showing that around 2 % were cytotoxic T cells (CD8<sup>+</sup>), 26 % were T helper cells (CD4<sup>+</sup>) and around 13 % were CD4<sup>+</sup>/CD8<sup>+</sup> double-positive cells. Within the pop-

ulation of T helper cells there were two more subpopulations to differentiate: naive CD4<sup>+</sup> T cells (CD45RA<sup>+</sup>/CD45RO<sup>-</sup>) and memory CD4<sup>+</sup> T cells (CD45RO<sup>+</sup>). Naive CD4<sup>+</sup> T cells represent about 6 % and memory T cells around 15 % of the c16orf54<sup>+</sup> cell population. The remaining cell population (11 %) of CD3<sup>-</sup>/CD19<sup>-</sup>/CD56<sup>-</sup>/CD4<sup>+</sup>/CD45RA<sup>+</sup> has not been identified so far.

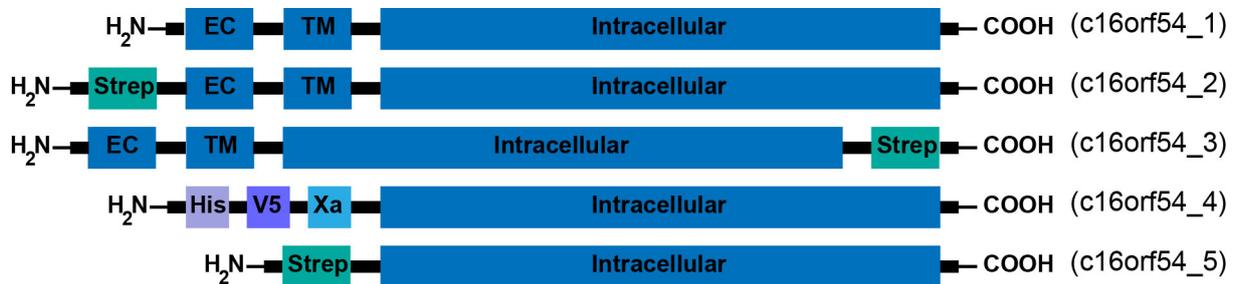


**Figure 29: Flow cytometry staining of human PBMCs – gated on c16orf54<sup>+</sup> cells.** Human PBMCs were stained with the purified and labeled antibody c16orf54 23H8\*-PE-TexasRed and specific antibodies against common lymphocyte cell surface markers (CD3: T cells, CD19: B cells, CD56: NK cells, CD4: CD4<sup>+</sup> T helper cell, CD8: cytotoxic T cell, CD45RA: naive phenotype, CD45RO: memory phenotype). The living lymphocyte population was first gated on c16orf54<sup>+</sup> cells and starting from this population (15.5 +/- 0.1 %) the percentage of lymphocyte subpopulations among the c16orf54<sup>+</sup> cells was examined. The distribution chart shows the percentages of the respective mother cell populations (n=3, mean +/- SEM). (n=3, flow cytometry plots shown for one representative donor)

### 2.3 Suitability screening of anti-human c16orf54 antibody (hybridoma supernatants) for Western blot

To investigate if the anti-human c16orf54 antibody is suitable for Western blot and immunoprecipitation, more control tools needed to be generated. This included the cloning of expression vectors carrying the c16orf54 coding sequence (Fig. 30). Three different full-length constructs were designed for the expression in HEK-293 cells, one without any change to the natural occurring form (\_1), one carrying an extracellular Strep-tag (\_2) and one carrying an intracellular Strep-tag (\_3). Two more constructs, considered for the expression in

insect cells, consisted only of the intracellular region combined with His- and V5-tag and a factor Xa cleavage site (\_4) or combined with a Strep-tag (\_5). The sequences were confirmed by sequencing.

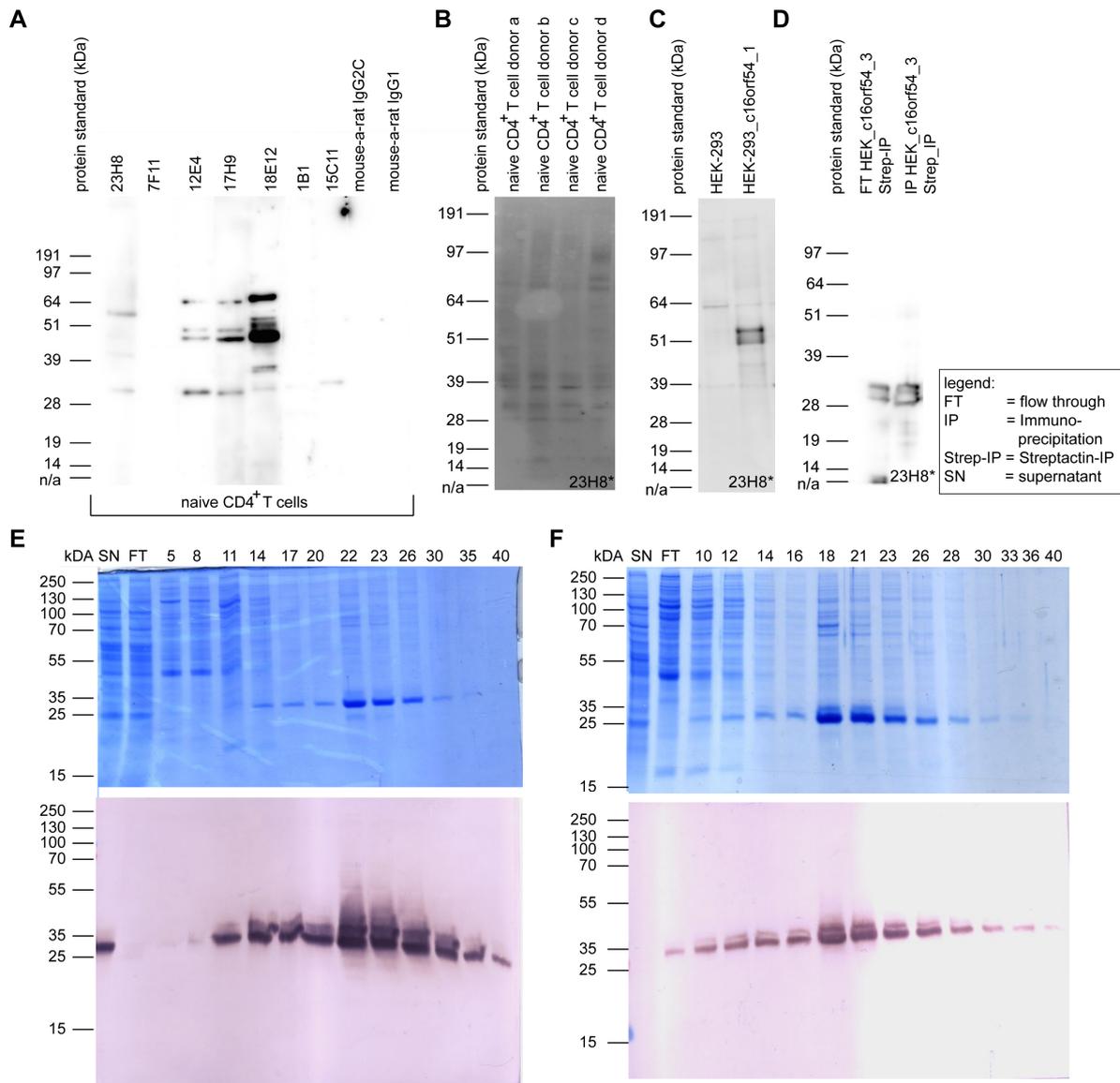


**Figure 30: Overview of the constructs for recombinant production of c16orf54.** Summary of constructs for recombinant expression of c16orf54 in mammalian cell lines (c16orf54\_1 – c16orf54\_3) and insect cells (c16orf54\_4, c16orf54\_5). (protein domains: EC=extracellular domain, TM=transmembrane region, Strep=Strep-tag, His=His-tag, V5=V5-tag, Xa=factor Xa cleavage site)

The Western blot suitability of the generated hybridoma supernatants with specificity for human c16orf54 was tested on samples of isolated human naive CD4<sup>+</sup> T cells (Fig. 31A). Used were the hybridoma supernatants, which already showed suitability for the use in flow cytometry stainings. The supernatant of clone 23H8 showed 3 bands, 7F11, 1B1, 27G3 did not show any signal, 12E4 and 17H9 both detected four bands, 18E12 showed 6 different bands and 15C11 one band. After consulting the IMI of the Helmholtz Center Munich, clone 23H8 was chosen as suitable antibody for Western blot application. Fig. 31B depicts the staining of a Western blot loaded with lysates of naive CD4<sup>+</sup> T cells of four different donors. The use of the purified antibody 23H8\* leads to the detection of similar bands as the unpurified hybridoma supernatant, but additionally detects bands in the higher molecular weight range. Therefore, 23H8\* was tested on a blot loaded with cell lysates of untransfected HEK-293 cells and HEK-293 cells overexpressing c16orf54 (Fig. 31C). This experiment indicated that the two bands with a molecular weight of around 51 kDa seem to be specific for c16orf54 under non-reducing conditions.

Fig. 31D shows a blot, which was loaded with fractions of an immunoprecipitation experiment (“flow-through” (FT) fraction, “immunoprecipitated” (IP) fraction). In this immunoprecipitation experiment, lysates of HEK-293 cells overexpressing the intracellular Strep-tagged c16orf54 (c16orf54\_3) were incubated with MagStrep beads to precipitate the protein of interest. The elution from the beads was done under reducing conditions. Again, two distinct bands of around 30 kDa (reducing conditions) were detected by antibody 23H8\* in both fractions, indicating that the IP with MagStrep beads was working in general (Fig. 31D). It can be concluded that the recombinant c16orf54 is detected by 23H8\* with two bands of around 51 kDa under non-reducing conditions and of around 30 kDa under reducing conditions. The immunoprecipitation setting was also carried out with lysates of HEK-293 cells expressing

the construct carrying the extracellular Strep-tag (c16orf54\_2), but in this case, antibody 23H8\* could not detect any band on the membrane (not shown here).

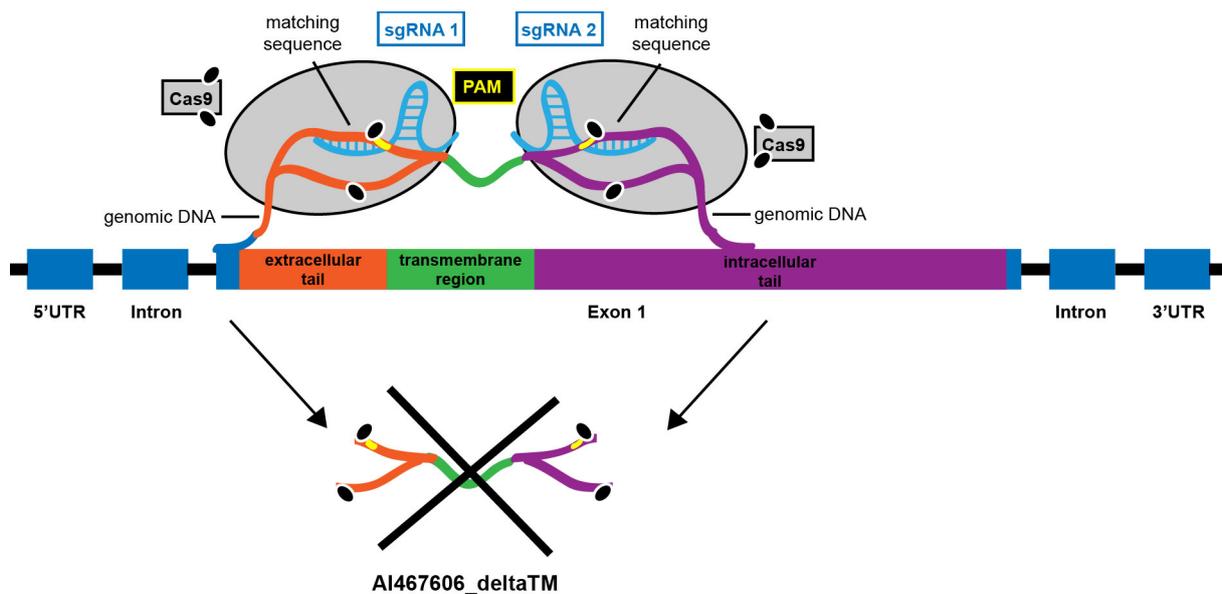


**Figure 31: Western blot analysis of cell lysates of human naive CD4<sup>+</sup> T cells using monoclonal antibodies with specificity for c16orf54 and validation of c16orf54\_4 and c16orf54\_5 expression and purification.** Hybridoma supernatants already positively tested in flow cytometry application were chosen to perform Western blot analysis. A) Cell lysates of human naive CD4<sup>+</sup> T cells of one donor were subjected to gel electrophoresis (same amount in all 11 lanes) and blotted. The membranes were probed with the respective hybridoma clone supernatants or antibodies against rat IgG subtypes as control. B) Cell lysates of human naive CD4<sup>+</sup> T cells of four different random donors were subjected to gel electrophoresis (same amount in all 9 lanes; reducing conditions) and blotted. The membrane was probed with the purified anti-c16orf54 antibody 23H8\* as primary antibody. C) Cell lysates of HEK-293 cells and HEK-293 cells overexpressing c16orf54 (HEK-293\_c16orf54\_1) were subjected to gel electrophoresis and blotted (non-reducing conditions). The membrane was probed with the purified anti-c16orf54 antibody 23H8\* as primary antibody. D) Immunoprecipitation with MagStrep beads was performed with cell lysates of HEK-293 expressing c16orf54\_3. The flow-through (FT) as well as the immunoprecipitated (IP) fractions were loaded on the gel, blotted and probed with anti-c16orf54 23H8\*. E) Coomassie staining of the different purified fractions of c16orf54\_4 with a molecular weight of around 25 kDa and a Western blot with the same fractions, probed with anti-V5-epitope antibody F) Coomassie staining of the different purified fractions of c16orf54\_5 with a molecular weight of around 25 kDa and a Western blot with the same fractions, probed with Strep-Tactin-AP conjugate.

Fig. 31E and F depict the Coomassie staining of the different purified fractions of c16orf54\_4 and c16orf54\_5 expressed in insect cells, which consist of the intracellular region of c16orf54 and varying tags. The purified fractions were additionally run on a second gel, blotted and detected using an antibody against the V5-epitope (Fig. 31E) in the case of c16orf54\_4 and a Strep-Tactin-AP conjugate (Fig. 31F) in the case of c16orf54\_5. The resulting bands confirmed the successful recombinant expression and purification of the c16orf54 constructs.

## 2.4 Strategy and proof of the *in vitro* functionality of sgRNAs for CRISPR/CAS-mediated knockout mice

To examine a potential *in vivo* function of c16orf54, a strategy to create a knockout mouse for AI467606, which is the murine homolog of c16orf54, was set up. To create a knockout mouse, the CRISPR/CAS system, in this case consisting of two sgRNAs targeting the flanking regions of the transmembrane coding domain of AI467606, was used (Fig. 32).

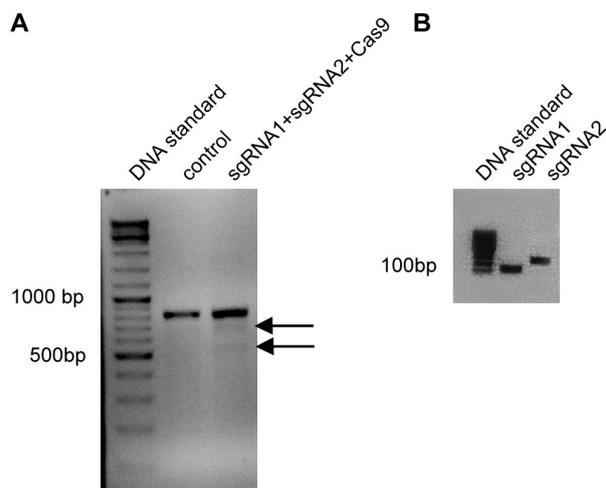


**Figure 32: Strategy for the generation of CRISPR/CAS mediated knockout mice.** This scheme shows the one-exon structure of AI467606, the murine homologue of the human c16orf54. The coding regions for the extracellular tail, the transmembrane region and the intracellular domain are color coded and named. To effectively generate a knockout of this region, PAM (protospacer adjacent motif) sequences (shown in yellow) flanking the transmembrane region on both sides were selected to clone sgRNAs targeting them. If both sgRNAs are functional, they should guide the Cas9 (DNA endonuclease enzyme) to the correct matching DNA sequence. Then, the Cas9 introduces double-strand breaks on both sides of the coding region of the transmembrane domain, including a small part of the extracellular- and intracellular tail. The repair mechanisms (non-homologous end joining or homologous recombination) of the cell promote the new connection of the two loose DNA ends without the targeted sequence, therefore creating a knockout mutation in the AI467606 gene.

After the cloning procedure of the constructs, they were tested *in vitro* on the murine cell line Neuro-2A. DNA was extracted from Neuro-2A cells, which were transfected with

constructs carrying sgRNA1, sgRNA2, Cas9 and only Cas9 as a control. Primers spanning the sgRNA-targeted region were used for PCR and the product was gel-purified. The purified PCR product was then incubated to let the DNA strands hybridize with each other. To find out if the hybridized strands show a mismatch-loop, which would be the case after a successful sgRNA targeting and Cas9 cleavage, the DNA was digested with T7 endonuclease. Fig. 33A shows the results of the T7 endonuclease incubation, confirming that a mismatch was detected by the enzyme only in the samples which were prepared from the cells transfected with the constructs carrying sgRNAs and Cas9, but not in the sample which was only transfected with Cas9. This reaction proofed the functionality of the sgRNAs *in vitro*.

For the preparation of the sgRNAs for injection into a zygote, the sgRNAs need to be transcribed *in vitro*. For this purpose T7\_PCR templates of the sgRNA1 and sgRNA2 were generated. Fig. 33B shows an agarose gel with bands at the expected sizes of approximately 133 bp for sgRNA1 and 275 bp for sgRNA2 and therefore depicts the proof of the successful generation of the T7\_PCR templates. The sgRNAs were then *in vitro* transcribed from the T7-PCR templates in a next step and purified for the injection into a zygote together with the Cas9 mRNA. A cooperation partner from the Institute of Developmental Genetics from the Helmholtz Center Munich performed the injection.



**Figure 33: Proof of sgRNA functionality and size-confirmation of T7-PCR templates for *in vitro* transcription.**

A) Neuro2A cells were transfected with only Cas9 (control) or with vectors carrying the sgRNA1, sgRNA2 and Cas9. T7 endonuclease did only detect a mismatch in the samples, which were transfected with Cas9 and both sgNAs. B) The size of the T7\_PCR templates for *in vitro* transcription of sgRNA1 and sgRNA2 was checked on an agarose gel.

28 mice were born, DNA was isolated for genotyping and sequencing revealed that none of these mice showed a CRISPR/CAS mediated deletion in the target locus of AI467606.

## VII. Discussion

Naive CD4<sup>+</sup> T cells form the basis of the immunologic memory, they possess the potential to differentiate into various kinds of T helper cell subsets and shape the actions of the adaptive immune system. Signals from cells of the innate immune system are recognized, processed in the T cell signaling network and lead to further signaling events within the immune system or other parts of the body, such as epithelial layers. An obvious key feature within this signal processing network is the composition of the cell surface molecules of the T cells, consisting of proteins, which function as receptors, transporters and adapters. Since it is known that the early signals, which lead to the activation and polarization of a T cell, essentially influence the first role and effector functions of the T cell, it is of high interest to gain insights into the early activation process as precisely as possible. The focus of the presented work is to show an in depth characterization of the cell surface of the naive CD4<sup>+</sup> T cell and to describe the changes happening in this compartment during the early time window of activation. Another aim was to identify cell surface proteins, which were not described in relation to T cells before and which therefore might be of interest as targets for the development of novel therapeutic approaches for various T cell mediated diseases. To be able to cover a broad T cell activation profile, the activation of the important cell surface receptors CD3 and CD28 within the T cell receptor complex, was utilized to mimic T cell receptor engagement.

### **1. Generation of a cell surface atlas of human naive and activated CD4<sup>+</sup> T cells – technical advantages and limitations**

In this study a comprehensive surface atlas describing proteins located on the surface of human naive and activated CD4<sup>+</sup> T cells is presented. The combination of omics approaches and high-throughput screening methods allowed an in depth characterization of the cell surface protein composition of human naive and anti-CD3/anti-CD28 activated CD4<sup>+</sup> T cells. Two proteomic and one transcriptomic technique were applied to link the transcriptome level to the proteome level. One mass spectrometry- and one flow cytometry-based approach led to the identification of 229 cell surface proteins, representing an experimentally verified pool of proteins which are present on the surface of human naive and/or activated CD4<sup>+</sup> T cells. To complement and extend the proteomic results on transcriptome level, a genome-wide microarray was linked to bioinformatics analysis. This analysis included the search for subcellular localization annotation based on experimental evidence or the prediction of sub-

cellular localization to identify transcripts, which encode cell surface proteins. Indeed, in the naive and/or shortly activated CD4<sup>+</sup> T cell (3h), 927 genes are expressed which code for proteins localized on the cell surface, 101 of them are also represented in one or both proteomic datasets (Table 2).

The first proteomic dataset is the result of the non-targeted, label-free PAL-qLC-MS/MS technique, which is based on the oxidation of sugar residues attached to glycosylated cell surface proteins. It is known that an oxidation treatment, like it is necessary for the PAL-qLC-MS/MS technique, could harm the cells by e.g. inducing cell death or introducing artificial changes in protein expression. This affects especially sensitive primary living cells, as it was seen by flow cytometry analysis during the establishment phase of the technique. But as the focus of this project was the examination of freshly isolated, primary human naive and activated CD4<sup>+</sup> T cells, the previously described PAL technique<sup>156</sup> had to be adapted for the use of these cells. To guarantee low side effects due to oxidation treatment, different NaIO<sub>4</sub> concentrations were tested. The results showed that mild oxidation conditions cause death of a certain percentage of cells. But still, the expression of selected cell surface proteins (such as CD69), analyzed via flow cytometry, showed similar results for the detected signal intensities when comparing untreated and 1 mM NaIO<sub>4</sub> treated cells. This means that under mild oxidizing conditions it is also possible to detect cell surface protein expression that is similar to the expression on cells, which were not treated with oxidizing agents. This was additionally monitored during the data acquisition via PAL-qLC-MS/MS of the samples (n=4) for the surface atlas and proved that the PAL-qLC-MS/MS technique leads to results, which are comparable to the results measured via flow cytometry. Another important aspect of the PAL-qLC-MS/MS data acquisition for the surface atlas was to analyze if the four different donors showed comparable results to combine them for further analyses. A Principal Component Analysis (PCA) of the acquired PAL-qLC-MS/MS data for the anti-CD3/anti-CD28 stimulation time course proved this to be true.

To evaluate the proteins identified via mass spectrometry and to generate a high confident cell surface dataset, a strict decision tree was built. Proteins, which passed the decision tree, were evaluated in regards to the number of identified peptides in the Trypsin and PNGase F fractions as well as the localization annotation of the proteins in the UniProt knowledge base.<sup>162</sup> By following this decision tree, 173 glycoproteins were identified on the surface of naive and/or activated CD4<sup>+</sup> T cells with high confidence. Within this set of 173 cell surface proteins, 54 proteins were identified in the MS analysis via only one peptide. This could indicate low expression or inefficient cleavage by the enzymatic digestion during sample preparation. One might also speculate that these one-peptide hits could be weaker candidates for the inclusion into the cell surface atlas. But the MS spectra of all one-peptide hits were manually inspected and could be declared as truly identified.

Unsupervised clustering gave an insight into the different expression patterns, which were detected via PAL-qLC-MS/MS during the stimulation time course. The analysis revealed that three distinct expression profiles could be specified within the dataset and all of them are characterized by specific immunology-associated Gene Ontology terms. Cluster 1 and 2 (38 % of the proteins), showed a fast downregulation during the first hours of T cell activation. Lck, CD3, CD4, CD28 and CD44 are cell surface proteins, contained in one of these two clusters, known to be present in the immunological synapse.<sup>185-189</sup> The fast turnover of cell surface proteins is in general an important feature, which enables cells to react very early after receiving any kind of stimulus, a fact that is especially important for proteins located in the immunological synapse. In contrast to that, Cluster 3 is composed of proteins, which can be described as late activation markers such as SEMA7A and CD274. They play an important role as negative regulators within the T cell response, which prevent overshooting reactions by terminating the T cell activation phase. This was also shown to be an important process for TCR signal termination, characterized by the clearance or internalization of the engaged T cell receptor as negative regulation of T cell activation.<sup>190</sup>

In addition to the classification of the cell surface glycoproteins into the different expression clusters, another aim was to assess if it is possible to describe the different early activation steps of the T cells by specific markers. By comparing the ten highest expressed cell surface proteins at all examined time points (0, 3, 6, 12, 24 and 48h) it became clear that this is not possible for the stimulation with anti-CD3/anti-CD28, as these ten highest expressed proteins were mostly the same at all time points. Interestingly, this analysis also showed remarkable higher protein abundance at 48h for the ten highest expressed proteins, compared to the other time points. For the *in vivo* situation, compared to the presented *in vitro* results, it was shown before by *in vivo* live imaging, that the T cells in the lymph node exhibited only short contacts to DCs around 48h after activation.<sup>191-193</sup> This might allow to speculate that the activation itself could be completed around this time point, proliferation and expansion are taking place, which require high amounts of energy involving many cell surface molecules for communication and transmembrane transport of supplements.

An extensive literature search revealed that 86 % of the 173 cell surface glycoproteins, including the ten highest expressed proteins at all stimulation time points, are associated with T cell biology or at least known to be involved in the biology of another cell type of the immune system. This is a strong confirmation for the robustness of the PAL-qLC-MS/MS approach. Interestingly, this literature search also revealed that 24 of these proteins were not co-cited with T cell biology at the time of the literature analysis. This situation changed when Bausch-Fluck *et al.* recently published a compendium, called the cell surface protein atlas (CSPA), which is a compilation of mass spectrometry derived datasets.<sup>151</sup> This atlas contains entities for 78 different human and murine cell types, including a T cell population

specified as CD4<sup>+</sup>/CD25<sup>-</sup>. This mixed population, containing non-activated naive as well as memory T cells, clearly differs from the pure naive CD4<sup>+</sup>/CD45RA<sup>+</sup>/CD45RO<sup>-</sup> T cell population, which was investigated in the presented study, and in addition Bausch-Fluck *et al.* did not focus on the activation process of these cells. Of note is, that 84 proteins, which were identified in the presented study by the PAL-qLC-MS/MS approach, were confirmed to be present in the Bausch-Fluck *et al.* dataset for CD4<sup>+</sup>/CD25<sup>-</sup> T cells, delineating that these cell surface proteins are most likely not specific for the naive, activated or memory type of CD4<sup>+</sup> T cell. Interestingly, the Bausch-Fluck *et al.* dataset for CD4<sup>+</sup>/CD25<sup>-</sup> T cells also included nine markers (ECE1, NPTN, RNF149, SLC1A4, SLC4A7, SLC5A3, SLC7A1, SYPL1, TMEM2) of these 24 cell surface proteins, which were not related to T cell biology so far.

Recently, Mitchell *et al.* published a multi-omic analysis (genome, methylome, transcriptome, miRNAome, proteome, phosphoproteome) of human naive and resting memory CD4<sup>+</sup> T cells of one blood donor.<sup>194</sup> They extracted potential cell surface proteins from their MS-derived dataset, using a stricter version of the bioinformatics strategy of the presented study (PAL-qLC-MS/MS), applied the same stricter filter to the PAL-qLC-MS/MS dataset and concluded that the datasets do have an overlap of 104 cell surface proteins. Additionally, they stated that 42 proteins were only detected by PAL-qLC-MS/MS and 1062 proteins were exclusively present in their dataset. This huge difference in the number of identified proteins is partly derived from different preparation strategies. The presented study applied a technique targeting only glycosylated cell surface proteins and the study of Mitchell *et al.* analyzed a size-fractionated whole cell lysate without focusing on the extraction of cell surface proteins during sample preparation. However, Mitchell *et al.* did not state within their manuscript, that the presented study used a labeling and purification method, which is proving that the cell surface proteins identified via MS are expressed on the cell surface at the time of sample preparation. In contrast, Mitchell *et al.* identified proteins in a whole cell lysate and extracted cell surface proteins only on the basis of bioinformatics annotation of subcellular localization. Proteins, which Mitchell *et al.* stated as cell surface proteins, might be localized intracellular at the time of sample preparation. A protein, which is intended to be expressed on the cell surface, might still be intracellular, as it could just have been released from the Golgi apparatus or the Endoplasmic reticulum during protein synthesis and post-translational modification and could not certainly be identified as a cell surface protein within a whole cell lysate.<sup>195,196</sup> It is also possible that some of these cell surface proteins are stored inside vesicles at the time of sample preparation. Of note is, that the surface dataset of Mitchell *et al.* introduces a large number of additional proteins, which might also be expressed on the surface of naive CD4<sup>+</sup> T cells. However, the localization of the proteins on the cell surface should be proven with a different proteomic technique, such as extracellular antigen staining coupled to flow cytometry analysis. Nevertheless, there will be surface

proteins, identified by Mitchell *et al.*, which are expressed on the surface of naive CD4<sup>+</sup> T cells, not identified by the presented PAL-qLC-MS/MS dataset. But this will be mainly due to the fact that the method used here is only able to detect glycosylated cell surface proteins and the method used by Mitchell *et al.* is able to detect proteins localized in all cellular compartments.

Currently, there are many techniques available, which are able to investigate cells on a single-cell level, such as single-cell PCR, flow cytometry or mass cytometry. The PAL-qLC-MS/MS technique results in datasets, which show the results for a pool of cells ( $8 \times 10^6$  cells) during the activation time course. The investigated population of naive CD4<sup>+</sup> T cells (0h) is characterized by the expression of CD3, CD4 and CD45RA, no expression of CD45RO<sup>197</sup> and the population also harbors naive CD4<sup>+</sup> T cells which are recent thymic emigrants (RTE).<sup>198</sup> Additionally, the current study showed that not all T cells, which were activated for 17h (Fig. 6B), simultaneously expressed the activation marker CD69 (Fig. 6B). This demonstrates that there are T cells, which are capable of transforming into the activated state at an early time point and other T cells which are reacting later to the stimulus of anti-CD3/anti-CD28, as it was also shown before.<sup>199</sup> The starting time point of the CD69 expression separates these two differentially reacting T cell populations, but there might also be other markers, which are characteristic for the early- and late-type reacting T cell. The PAL-qLC-MS/MS technique, as stated before, detects the sum of changes within the activated T cell population at the different time points and is not able to dissect the changes on single cell level. But as the PAL-qLC-MS/MS technique is a non-targeted approach, only relying on the glycosylation of the cell surface proteins, it was possible to detect proteins with this technique, which were not related to T cell biology before. These findings are initiating the start to specifically investigate these “novel” described proteins on single cell level during the T cell activation process, by the staining with a monoclonal antibody coupled to flow cytometry analysis for example. Maybe, it is possible to then describe the different activated T cell populations more detailed to better understand the early differences and gain more insights into the activation/stimulation process in general.

To extend and confirm the PAL-qLC-MS/MS dataset of cell surface glycoproteins by another proteomic technique, a high-throughput flow cytometry-based surface antigen screening was conducted, which includes the possibility to investigate the T cells on a single-cell level. The expression of 332 cell surface proteins was examined by single stainings with monoclonal antibodies. This analysis could confirm the expression of 67 proteins, which were identified by PAL-qLC-MS/MS before and expanded the proteomic cell surface dataset about further 56 surface proteins, which were only identified by the flow cytometry screening. By closer examination and comparison of the proteins identified by PAL-qLC-MS/MS and/or the flow cytometry screening (samples of time points, which were measured with both tech-

niques), some divergent detection trends were noticeable at 0 and 3h. In most of the cases, these differences originate from a positive detection by PAL-qLC-MS/MS technique and a flow cytometry signal, which was under the strict IgG control threshold. The proteins, which are affected by these divergent detection trends, are mostly shared between the 0 and 3h stimulation time point. This observation might be a result of the threshold set for the evaluation of the flow cytometry results, which was very strict, potentially resulting in a false negative assumption. Nevertheless, these proteins were included in the cell surface atlas, because they were detected via PAL-qLC-MS/MS. The combination of these results led to a dataset containing 229 cell surface proteins in total, identified on protein level on naive and/or activated CD4<sup>+</sup> T cells.

Indeed, the outcome of the two different proteomic approaches on the one hand relies on the glycosylation of the proteins (PAL-qLC-MS/MS) and on the other hand is dependent on the availability of an antibody (flow cytometry screening). To partially compensate this obstacle, a genome-wide transcriptomic analysis coupled to bioinformatics was conducted to estimate the potential number of cell surface proteins, which are expressed on mRNA level. Briefly, transcripts expressed in naive and/or 3h activated CD4<sup>+</sup> T cells were analyzed concerning the subcellular localization of the corresponding proteins. Subcellular protein localization annotation (UniProtKB) as "cell surface" or "plasma membrane" was considered, when the annotation was experimentally verified. If this was not the case, the subcellular localization was predicted by using the recently published LocTree3 (accuracy 80 %), substantiated with the prediction of at least one transmembrane helix. The selection criteria for the identification of transcripts, which encode cell surface proteins, were stricter than the analysis of the PAL-qLC-MS/MS dataset to reduce the number of false positives. Because in contrast to this analysis, the PAL-qLC-MS/MS approach and also the flow cytometry screening provided direct experimental evidence for the localization of the protein on the cell surface. The outcome of this *in silico* analysis and prediction of cell surface protein encoding transcripts, revealed 927 potentially expressed cell surface proteins. This number might seem low compared to the study of da Cunha *et al.*<sup>200</sup>, who identified 3700 genes, believed to code for human cell surface proteins. The strategy of da Cunha *et al.* to identify cell surface coding transcripts, is based on the whole sequence of the human genome (National Center for Biotechnology information build 36.1) coupled to bioinformatics analysis tools. The number of predicted cell surface coding transcripts, which was identified in the presented study, is in line with other estimations supposing that approximately 26 % of the human genes code for cell surface proteins.<sup>201</sup> However, the reduced number of transcripts in the transcriptomic cell surface dataset presented in the current study appears to be realistic, because this dataset only focused on naive and activated CD4<sup>+</sup> T cells and the selection criteria which were utilized were stricter than in the study of da Cunha *et al.*

To get more detailed information about the 927 cell surface protein coding transcripts, a GO term enrichment analysis was performed. The resulting GO terms point to biological functions, which are generally fulfilled by plasma membrane proteins such as signal transduction or cell adhesion. However, this analysis is not strongly associated with an immunological function of the transcripts in contrast to the GO term analysis of the cell surface proteins identified via PAL-qLC-MS/MS. To narrow down this group of 927 cell surface protein coding transcripts, they were examined concerning differential expression in naive and 3h activated CD4<sup>+</sup> T cells. A significantly different expression in the two examined states of the CD4<sup>+</sup> T cell (0, 3h) was shown for 141 of the transcripts. The ten genes with the highest expression difference are *SEMA7A*, *FASLG*, *CLIC4*, *FFAR3*, *CD200*, *TMEM88*, *TNF*, *SLC7A5*, *TNFSF14* and *CD69*. For most of these ten genes, but also for most others of the 141 transcripts, a role in T cell activation and differentiation is described, a fact that acknowledges the robustness of the cell surface coding transcript filtering approach. Besides the targets, which are already described as relevant for T cell biology, also other genes such as *FFAR3* (free fatty acid receptor 3) appeared in this analysis. *FFAR3* was recently described in a murine study as crucial sensor on dendritic cell precursors for short-chain fatty acid transport<sup>202</sup>, and is among the top ten differentially expressed transcripts in shortly activated human CD4<sup>+</sup> T cells. Interesting candidates like *FFAR3* need to be further investigated concerning their role in the context of T cell activation.

The strength of the transcriptional cell surface dataset can be demonstrated by comparing the outcome of the proteomic approaches with the transcriptomic results. Indeed, under the strict bioinformatics selection criteria of the *in silico* analysis of the transcriptomic results, 53 % of the proteins (PAL-qLC-MS/MS and flow cytometry screen) would have been detected on transcriptional level and also identified as cell surface protein coding transcripts. When loosening the strict criteria, meaning the inclusion of proteins annotated as “membrane” (without further information on localization) and considering the putative dataset of 248 additional transcripts with the subcellular annotation as “lipid anchor” or “peripheral”, this overlap could be increased to 83 %. On the one hand, these numbers illustrate the limitations of trusting only transcriptional data combined with bioinformatics, as 108 of the detected proteins would have not been identified in the strict transcriptional dataset. This emphasizes on the necessity of non-targeted proteomic approaches. However, on the other hand, the cell surface protein coding transcripts of the strict dataset (without the ones overlapping the proteomic targets) combined with the putative dataset (n=1074), demonstrate a large pool of further interesting targets, which of course would first need to be verified on protein level. The reason why many of the cell surface protein coding transcripts were not detected on the protein level by one of the two proteomic approaches, might be explained by the fact that not all existing transcripts are translated into proteins at every time.

Moreover, also posttranslational modification, translocation and degradation of proteins need to be considered, as well as the limitation of the PAL-qLC-MS/MS of only detecting glycosylated cell surface proteins. The bioinformatics analysis (UniProtKB) of the strict transcriptomic cell surface dataset suggested a N- or O-glycosylation site for 70 % of the proteins encoded by these transcripts (Supplemental Table S3). Therefore, the PAL-qLC-MS/MS approach fails to detect approximately 30 % of the cell surface proteins on the naive and activated CD4<sup>+</sup> T cell. However, the combination of the PAL-qLC-MS/MS technique with the flow cytometry screening and the genome-wide transcriptomic approach is at least partially able to overcome this limitation. These two additional techniques extend the number of cell surface targets in the cell surface atlas and provide a pool of potential further interesting candidates, which can be evaluated in future studies.

Taken together, the comprehensive cell surface atlas, generated via transcriptomic as well as proteomic techniques, presents a reference book for the surface profile of human naive and activated CD4<sup>+</sup> T cells. It broadly describes the composition of the surface proteins, giving the possibility to deeply monitor the events happening on the surface during the early time window of T cell activation.

## **2. Potential of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells for future studies**

How to choose interesting candidate targets for further investigations from large and mostly descriptive omics datasets is a recurrent question.<sup>151</sup> In this work, such a rich pool of targets, shown to be expressed on the cell surface of naive and/or activated CD4<sup>+</sup> T cells, is presented. The strategy of this work to decide which of the targets are of value to be selected as candidates for further investigations, are based on three ideas. One of the ideas was to combine the findings on transcriptomic and proteomic level and choose targets, which were identified by both approaches and which were differentially expressed within the stimulation time course. This analysis defined 32 out of 229 cell surface proteins, which showed a significant difference in expression in at least one of the used techniques at a minimum of one stimulation time point. Many of these 32 proteins have a defined role in the context of T cell biology, like CD69, a known T cell activation marker<sup>203</sup>, and CD108<sup>204</sup>, which is known to be involved in the modulation of inflammation and the T cell mediated immune response. 21 proteins of this group are already affiliated to the Cluster of Differentiation (CD), meaning that they have a described role within the immune system in general or a specified function for T cell biology like CD25, the IL-2 receptor. A second group, which is numerically well represented in this analysis, is the protein family of solute carrier (SLC) transporters. One

representative of this group is SLC7A5, the central large neutral amino acid transporter, which is also known as CD98LC. The importance of this transporter for T cells, regarding the differentiation into effector subsets and clonal expansion upon antigenic T cell receptor stimulation, was shown in a murine study.<sup>205</sup> In the presented study SLC7A5 was among the top ten significantly different expressed targets on RNA level and was present in the combined analysis of transcriptomic and proteomic results of human naive CD4<sup>+</sup> T cells. This transporter shows a strong upregulation upon anti-CD3/anti-CD28 stimulation in the PAL-qLC-MS/MS Cluster 3. Additionally, this was also recently shown for CD4<sup>+</sup>/CD25<sup>-</sup> T cells.<sup>206,207</sup> Another interesting target, which resulted from the analysis of the combined omics datasets, is the protein EVI2A (ecotropic viral integration site 2A, gene name: *EVI2A*), also present in the PAL-qLC-MS/MS Cluster 3. The corresponding gene *EVI2A* is situated together with *EVI2B* in a long intron of the *NF1* (neurofibromin 1) gene, which was demonstrated to be deleted in some patients suffering from neurofibromatosis.<sup>208,209</sup> The relation of the *EVI2A* locus to the hematopoietic system was shown by the ability of the locus of its murine homologue *EVI2* (homologue of human *EVI2A*) to act as an integration site for the Murine Leukemia Virus. Upon infection with this virus and integration into this gene locus, it is possible that myeloid leukemia develops.<sup>210</sup> The protein EVI2A, expressed in the human system, is thought to be able to form a cell surface receptor as homo- or heterodimer with other proteins (information was obtained from UniProtKB). The protein encoded by *EVI2B*, one of the genes in close proximity to *EVI2A*, was also identified in the present study on transcriptional as well as protein level. The protein EVI2B, in contrast to EVI2A, was already affiliated to the Cluster of Differentiation (CD) as CD361, a new B cell marker, at the Human Leukocyte Differentiation Antigens (HLDA) Workshop in 2010.<sup>211</sup> But it was also shown that EVI2B is expressed on granulocytes, monocytes, NK cells and T cells, which implies that this protein might have a general role or function on different cells of the immune system. Nonetheless, neither EVI2B, nor EVI2A are functionally characterized so far, therefore further studies are very important to elucidate their function in general and especially their role in the context of T cell biology.

Nowadays, there are many possibilities to modify protein expression in the different stages between the transcription of a gene and the generation of the final functional posttranslational modified protein. These modifications can be achieved by numerous modern tools like RNA interference, recombinant molecules acting as agonists or antagonists, small molecules used as inhibitors, substrates or inducers, or therapeutic antibodies. According to the stage, which shall be targeted, the tool must be precisely selected. Sometimes studies rely on protein or transcript data only. However, as other studies before<sup>212-216</sup>, the study presented here, demonstrated very weak correlation of the expression level of transcript and protein data. It undermines the notion that relying on only one of these information levels might

lead to failure. The only point, which can be concluded for naive CD4<sup>+</sup> T cells by the presented correlation data is, that early transcriptomic expression levels (0 and 3h) are partly reflected in early protein expression levels (Spearman correlation coefficient 0.6), but not reflected or only very weakly reflected in late protein expression levels (Spearman correlation coefficient 0.3). This permits to draw the conclusion that transcriptomic as well as proteomic levels of the target of interest need to be investigated separately to get a clear and complete picture, like it was performed for the cell surface atlas of naive and activated CD4<sup>+</sup> T cells. This comprehensive knowledge is therefore essential to be able to select the correct tool for modifying the target of interest within the biological system.

Although, it is not very easy to choose candidates for further investigations from large omics datasets, Mirkowska *et al.*<sup>152</sup> previously demonstrated the high translational potential of such datasets, focusing on the description of cell surface proteins. The authors of this study were able to identify a new diagnostic candidate marker for leukemia by investigating xenografts from 19 patients with B cell precursor acute lymphoblastic leukemia by the analysis of cell surface markers. A second study by Strassberger *et al.*<sup>217</sup> focused on the analysis of the surface proteome of myeloid leukemia cell lines. They were able to identify a cell surface target, which exhibited differential expression between diseased and healthy cells. In addition, they managed to produce a monoclonal antibody equipped with tumor targeting properties *in vivo*, demonstrating the potential for the development of therapeutic strategies based on comprehensive cell surface proteome datasets. The second idea on how to identify interesting candidate targets in the presented cell surface atlas of naive and activated CD4<sup>+</sup> T cells for further investigations, was therefore of pragmatic origin. Thus, the focus of interest was the discovery of targets for drug development and application. As cell surface proteins are easy accessible and therefore valuable targets for drugs, all 229 cell surface proteins of the presented cell surface atlas were examined concerning their annotation as drug target. Regarding antibody-based therapeutic strategies, the presented cell surface atlas contains 7 proteins, which are targeted by approved therapeutic antibodies. CD52 and CD152 are targeted by antibodies in the context of cancer therapy, while CD3, CD11a, CD25, CD49d and CD126 are targeted by antibodies for the therapy of autoimmune diseases ([www.antibodysociety.org](http://www.antibodysociety.org), [www.immunologylink.com](http://www.immunologylink.com)). By considering a very recent publication on the current pipeline of therapeutic antibodies from the biopharmaceutical industry, it becomes clear that the cell surface atlas of naive and activated CD4<sup>+</sup> T cells includes more proteins, which are interesting targets for novel therapeutics. Three antibodies (targeting CD274, CD126 or CD4) are in phase 3 clinical trials, whereas antibodies against CD38 (already approved in the United States of America), CD26 and CD319 are already under regulatory review within the European Union.<sup>155</sup> Furthermore, the high translational potential of the cell surface atlas is also underlined by the results of a DrugBank database search.<sup>168</sup> As it

is presented in Table 2, 60 of the 229 identified cell surface proteins are targets for approved (n=51), or experimentally validated/currently under investigation drugs (n=9). A prominent group within these targets for approved drugs is the solute carrier (SLC) transporter protein family. This protein family is composed of transporters for all kinds of molecules, such as amino acids, ions, sugars, and ATP/ADP. On genome-wide transcriptome level, 220 SLC coding transcripts were detected in naive and/or activated CD4<sup>+</sup> T cells. The proteins encoded by 85 of these transcripts are annotated or predicted to have cell surface localization and, indeed, 28 proteins of this family were confirmed by the PAL-qLC-MS/MS technique to be expressed on the cell surface. Most of them are only low expressed on the naive CD4<sup>+</sup> T cell, but increase in expression during anti-CD3/anti-CD28 stimulation. In most cases, their highest abundance was detected at 48h of activation, defining them as interesting targets for a later stage of the activation process. Interestingly, 14 of these SLCs are annotated as targets for approved drugs (2 more are annotated as targets for approved drugs, but the drugs were withdrawn). These drugs, targeting the SLCs on activated CD4<sup>+</sup> T cells, can now be examined concerning their impact on the T cell fate. To investigate the effect of already approved drugs on proteins expressed on CD4<sup>+</sup> T cells, has a high translational potential for the development of therapeutic approaches for T cell mediated diseases. If a positive or negative effect on activation or even a subset-shifting effect on differentiation of the CD4<sup>+</sup> T cells can be achieved by these drugs, this could give rise to new ideas regarding vaccination or immunotherapeutic strategies.

The third idea on how to choose interesting targets from the cell surface atlas for further investigations is to analyze the currently available information about the detected proteins, as well as their association to T cell biology. As stated previously, an extensive literature, protein database and patent search revealed that 24 of the 229 detected proteins were not related to the biology of human naive CD4<sup>+</sup> T cells and their early activated state at the time of analysis. This group of proteins consists of members of the SLC family (SLC1A4, SLC25A3, SLC25A5, SLC4A7, SLC5A3, SLC5A6, SLC7A1), proteins with an annotated function on cells not belonging to the immune system (APOC3, ATP1A1, CALML5, DCD, ECE1, ERO1L, GOT2, GPA33, INHBB, NPTN, RNF149, SBSN, SMR3B, SYPL1) and proteins, lacking a functional characterization within the human system (transmembrane protein c16orf54, EVI2A, TMEM2). The expression of the membrane-anchored proteins (n=20) of this group was validated via qPCR. The availability of validated antibodies for Western blot techniques, allowed the additional confirmation of the expression of four of these targets (EVI2A, RNF149, NPTN, TMEM2) on protein level. A validation of more of these targets with an alternative protein-detecting technique is pending until more validated detection tools, such as monoclonal antibodies, will be available. At this stage, these 24 cell surface proteins could not be specified as new T cell markers. Their expression on protein level should additionally

be verified by another proteomic technique and the distribution on other PBMC subtypes needs to be investigated to be able to call them “T cell surface makers”. However, the status of these 24 proteins, which were not related to T cell biology before the presented study, changed when Bausch-Fluck *et al.*<sup>151</sup> published the cell surface protein atlas (CSPA). Within their dataset of CD4<sup>+</sup>/CD25<sup>-</sup> T cells, they verified the presence of nine (ECE1, NPTN, RNF149, SLC1A4, SLC4A7, SLC5A3, SLC7A1, SYPL1, TMEM2) of these 24 cell surface proteins. This additional dataset confirms the presence of these proteins on human T cells and emphasizes even more that there is a need to investigate their potential as novel drug targets on T cells. Moreover, functional studies of these proteins are an essential next step to find out in which way they might be relevant for T cell biology.

Taken together, the three strategies to reveal interesting candidates for further investigations, pointed to a recurring set of proteins. This set contains the very interesting members of the SLC family, which are promising candidates for the investigation of the effects of approved drugs, which are targeting them. But first, for some of them also a potential novel role in T cell biology needs to be investigated to collect more information. An interesting candidate for such investigations is also EVI2A, a protein, which was present in more than one of the target-choosing analyses. But also proteins like the transmembrane protein c16orf54, another protein without a known function, displays a promising target for further investigations.

### **3. c16orf54 – a novel described cell surface protein on cells of the immune system**

The transmembrane protein c16orf54 was discovered on the surface of human naive CD4<sup>+</sup> T cells during the PAL-qLC-MS/MS data acquisition. This protein was chosen as a first candidate for further investigations from the pool of the 24 cell surface proteins, which were not described in relation to human T cell biology at the time of data analysis. The only available information about this protein at the beginning of this work was a murine study by Ferreras *et al.*<sup>218</sup>, which showed that the murine homologue of c16orf54 is a target for the transcription factor RUNX1/AML1 and that it is expressed within the hematopoietic cell lineage. So, in general, there was only little information available about this protein within the human system. It is the product of the chromosomal open reading frame 54 on chromosome 16. Its predicted molecular weight is 25 kDa and it is supposed to be O-glycosylated<sup>219</sup> on its small extracellular domain, which comprises 31 amino acids. The transmembrane region consists of 21 amino acids, and the intracellular domain is 172 amino acids long and contains three potential phosphorylation sites, predicted by similarity to the murine and the

rat homologue.<sup>162</sup> A search for conserved domains or motifs within the amino acid sequence of this protein via an online search tool (Swiss-Prot)<sup>162</sup>, gave no results. To further characterize this protein, we aimed to develop specific monoclonal antibodies. Therefore, a monoclonal antibody directed against the extracellular domain was produced via the hybridoma-technique in cooperation with the Institute of Molecular Immunology of the Helmholtz Center Munich. The hybridoma supernatants were validated via flow cytometry and a peptide-competition staining was able to confirm the specificity of the selected purified antibody (23H8\*). Beneath the functionality in flow cytometry, the antibody 23H8\* was also tested for its functionality in Western blot. On samples of c16orf54 recombinantly producing HEK-293 cells, it could be shown that the antibody detects the protein in the range of 30 kDa under reducing and of 51 kDa under non-reducing conditions.

To get information about the distribution of the transmembrane protein c16orf54 on human PBMCs, the antibody 23H8\* was directly labeled for flow cytometry first. Interestingly, within the population of PBMCs, around 15.5 % of the cells are positive for c16orf54, the two main populations among them are T cells (around 43.5 %) and B cells (around 41.6 %). These results clearly indicate that c16orf54 is not exclusively expressed on T cells, as nearly the same amount of B cells expresses this marker. By examining specified subpopulations of the PBMCs, it could also be shown that c16orf54 is not expressed on the whole subpopulation of a distinct type of cells. Around 80 % of the B cells, 50 % of the NK cells, 78 % of the NKT cells, 17.8 % of the memory CD4<sup>+</sup> T cells and only 8 % of the naive CD4<sup>+</sup> T cells express c16orf54. This fact points to a potential general role of c16orf54 within the population of human PBMCs, as it is expressed on many of the different cell subsets. However, it seems to separate cells within a distinct subpopulation into c16orf54<sup>+</sup> and c16orf54<sup>-</sup> cells. During these investigations, a very recent study was published<sup>220</sup>, which identified c16orf54 (SAIL, surface antigen in leukemia) as an important target on leukemia patient samples via a mass spectrometry-based technique. Spectral counts from these MS results and flow cytometry results (self-made monoclonal antibody) indicated that c16orf54 is significantly higher expressed in samples of chronic lymphocytic leukemia (CLL) patients compared to acute myeloid leukemia (AML), multiple myeloma (MM), and bone marrow mononuclear cells (BMMC) as well as PBMCs of healthy donors. But the authors also confirmed that this protein is expressed on at least 10 % of healthy PBMCs, which was also shown in the presented study. By an RNA ISH (in situ hybridization) assay, Kim *et al.* investigated c16orf54 expression in normal tissue and B cell lymphoma samples. Regarding the normal tissue, they showed that c16orf54 is moderately expressed in bladder urothelium and lymphoid tissue (lymph node, spleen, tonsil and thymus), and only very low expressed in epithelial tissue (uterus, pancreas, gallbladder, cervix, esophagus) and not detectable in other types of epithelial tissue. Interestingly, by analyzing the samples of B cell lymphomas they could show c16orf54 expression in follicular

lymphoma and activated B cell (ABC) as well as germinal center B cell (GCB) type of diffuse large B cell lymphoma (DLBCL). As cancer cells of CLL patients are characterized by immune suppression ability<sup>221</sup>, and as the results of Kim *et al.* and of the presented study show that cancer cells of the CLL type<sup>220</sup> and only subsets of healthy PBMCs express c16orf54, it might cautiously be speculated that c16orf54 could be a marker related to immune response suppression. But of course, this needs to be proven by a solid experimental system. Efforts to obtain information about the function of c16orf54 by identifying interaction partners of this protein were undertaken by performing immunoprecipitation (IP) experiments, but were not successful, as the 23H8\* antibody seems to be not stable enough in buffers which are frequently used for IP experiments. Recently published datasets on high-throughput proteomic mapping of interaction networks in the human system by Huttlin *et al.*<sup>222</sup>, accessible via the database BioGrid<sup>223</sup>, propose IFNA17 (interferon alpha 17), PKP4 (plakophilin-4) and EFNB1 (ephrin-B1) as interaction partners for c16orf54. However this needs to be proven for the cells of the immune system. But still, this is a very important point to follow up on, as e.g. low doses of ephrin-B1 and ephrin-B2 were shown to be able to co-stimulate T cell proliferation induced by anti-CD3, and high doses were shown to block T cell proliferation in a murine study.<sup>224</sup>

To examine a potential *in vivo* function of c16orf54 and the importance of c16orf54 for the body, it is of great interest to switch to the murine system, which provides many opportunities such as knockout (KO) and disease models. AI467606 is the murine homologue of c16orf54 and, as stated before, was previously described to be regulated by the transcription factor RUNX1/AML1.<sup>218</sup> Ferreras *et al.* also demonstrated that AI467606 is already expressed during the onset of the development of the hematopoietic system. Additionally it is expressed on many cells of the hematopoietic lineage in adult mice. To follow up on this and to gain a deeper insight into the function and importance of the expression of AI467606, two strategies were followed. The first strategy was to buy sperm for the creation of a reporter/KO mouse (AI467606\_KOfirst, see scheme of the targeting vector in Appendix Figure A4), containing lacZ as reporter gene for histology and flow cytometry and a stop codon after the lacZ to be able to breed a total KO mouse. Moreover, the coding exon of AI467606 is flanked by loxP sites as acceptor regions for Cre-recombinase to be able to breed a conditional KO mouse in the future. The first generation of these mice (AI467606\_KOfirst) was born and genotyped (Appendix Figure A5), resulting in seven transgenic mice, which will first be used to follow AI467606 expression via the lacZ reporter in the murine development via histology and flow cytometry. This lacZ reporter will then also serve as positive control for the generated monoclonal antibody 13D2\*, which is specific for the murine homologue of c16orf54. The second strategy was to create a KO mouse line for AI467606 in-house, on the basis of collaboration with the Institute of Developmental

Genetics of the Helmholtz Center Munich. The currently fastest way to generate a KO mouse is via the CRISPR/CAS system.<sup>225</sup> Therefore, a strategy using two sgRNAs flanking the region, coding for the transmembrane domain of AI467606, was set up. If selected sgRNAs are functional, they will target the Cas9 nuclease to the sequence of the transmembrane region of AI467606 to introduce double-strand-breaks (DSB) on each transmembrane-region-flanking side. The cellular repair mechanisms would then fuse the loose ends together by non-homologous end joining or homologous recombination, resulting in a potential frame shift due to small insertions or deletions (indels). In the end, the protein is not expressed in the correct way or it is not expressed at all. The functionality of the sgRNAs was proven in an *in vitro* test using the murine cell line Neuro-2A. The sgRNAs were *in vitro* transcribed and purified for the injection into a zygote to produce the AI467606\_deltaTM\_CRKO mouse (TM=transmembrane). Within the offspring of the first microinjection, there are no mutant mice. This could be the result of low *in vivo* efficiency of the sgRNAs or a general technical problem during the microinjection process. The injection needs to be done again to find out if the sgRNAs will work in a second attempt. Although the successful *in vitro* test in the Neuro-2A cell line is a good predictor for the sgRNA activity, it is also possible that the same activity cannot be achieved in other cell lines or in mouse embryos. Thus, if a second microinjection will also not result in the generation of mutant mice, then the reporter mouse line AI467606\_KOfirst needs to be bred to a total KO, to gain functional insights into an AI467606 KO mouse model.

Taken together, c16orf54 was detected on naive CD4<sup>+</sup> T cells during the generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells. The establishment of investigational tools, such as the monoclonal antibodies and the recombinant expression systems, led to first results concerning the distribution of c16orf54 on human PBMCs, which indicated that c16orf54 is not a marker, which is specific for T cells, as it is expressed on many cell subsets within PBMCs. Together with very recent literature, showing that c16orf54 is frequently expressed on patient samples of different blood cancer types such as CLL<sup>220</sup>, this might allow the hypothesis that c16orf54 could be related to a certain immune suppression functionality of cells in the hematopoietic compartment and confirms the relevance of investigating this protein in more detail. Following up on three proposed interaction partners of c16orf54<sup>222</sup> will raise the opportunity to perform proof of concept experiments by using the novel generated investigational tools, leading to the possibility to embed c16orf54 into the cellular signaling network of cells of the immune system. The basis for further functional analyses of c16orf54 was built by the switch to the murine system and the generation of mutant mouse lines. One of these mouse lines, which is already established, will give the opportunity of a reporter system for c16orf54, a total knockout and a conditional knockout. First, the phenotyping of the reporter mouse line will provide insights into the distribution of c16orf54

---

within the murine system. In the future, one of the total KO models and the conditional KO model will help to gain insights into the importance of c16orf54 for the murine development. Furthermore, different disease models for immune system disorders can be applied to the mice to gain further functional insights into the regulation of the protein, but also to find out if c16orf54 plays any major role within immune diseases and might then be a promising structure to target for the development of novel therapeutic applications.

## References

- 1 Abbas, A. K., Lichtman, A. H. & Pillai, S. *Cellular and molecular immunology*. (2012).
- 2 Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464**, 1367-1370 (2010).
- 3 Mjosberg, J. M. *et al.* Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. *Nat Immunol* **12**, 1055-1062 (2011).
- 4 Janeway, C. A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216 (2002).
- 5 Boehm, T. & Swann, J. B. Origin and evolution of adaptive immunity. *Annu Rev Anim Biosci* **2**, 259-283 (2014).
- 6 Kanno, Y., Vahedi, G., Hirahara, K., Singleton, K. & O'Shea, J. J. Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. *Annu Rev Immunol* **30**, 707-731 (2012).
- 7 Zhu, J., Yamane, H. & Paul, W. E. Differentiation of effector CD4 T cell populations. *Annu Rev Immunol* **28**, 445-489, doi:10.1146/annurev-immunol-030409-101212 (2010).
- 8 Ciofani, M. & Zuniga-Pflucker, J. C. The thymus as an inductive site for T lymphopoiesis. *Annu Rev Cell Dev Biol* **23**, 463-493 (2007).
- 9 Koch, U. & Radtke, F. Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol* **27**, 539-562 (2011).
- 10 Dontje, W. *et al.* Delta-like1-induced Notch1 signaling regulates the human plasmacytoid dendritic cell versus T-cell lineage decision through control of GATA-3 and Spi-B. *Blood* **107**, 2446-2452 (2006).
- 11 Hsieh, C.-S., Lee, H.-M. & Lio, C.-W. J. Selection of regulatory T cells in the thymus. *Nature reviews Immunology* **12**, 157-167 (2012).
- 12 Fooksman, D. R. *et al.* Functional anatomy of T cell activation and synapse formation. *Annu Rev Immunol* **28**, 79-105 (2010).
- 13 Bachmann, M. F., Barner, M. & Kopf, M. CD2 sets quantitative thresholds in T cell activation. *J Exp Med* **190**, 1383-1392 (1999).
- 14 Singleton, K. *et al.* A large T cell invagination with CD2 enrichment resets receptor engagement in the immunological synapse. *Journal of immunology (Baltimore, Md : 1950)* **177**, 4402-4413 (2006).
- 15 Espagnolle, N. *et al.* CD2 and TCR synergize for the activation of phospholipase Cgamma1/calcium pathway at the immunological synapse. *Int Immunol* **19**, 239-248 (2007).
- 16 Boyman, O., Letourneau, S., Krieg, C. & Sprent, J. Homeostatic proliferation and survival of naive and memory T cells. *European journal of immunology* **39**, 2088-2094 (2009).
- 17 Brzostek, J., Gascoigne, N. R. J. & Rybakina, V. Cell Type-Specific Regulation of Immunological Synapse Dynamics by B7 Ligand Recognition. *Frontiers in immunology* **7**, 24 (2016).
- 18 Malherbe, L., Hausl, C., Teyton, L. & McHeyzer-Williams, M. G. Clonal selection of helper T cells is determined by an affinity threshold with no further skewing of TCR binding properties. *Immunity* **21**, 669-679 (2004).
- 19 Chu, H. H. *et al.* Positive selection optimizes the number and function of MHCII-restricted CD4+ T cell clones in the naive polyclonal repertoire. *Proc Natl Acad Sci U S A* **106**, 11241-11245 (2009).
- 20 Mosmann, T. R. & Coffman, R. L. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. *Annu Rev Immunol* **7**, 145-173, doi:10.1146/annurev.iy.07.040189.001045 (1989).
- 21 Biedermann, T., Rocken, M. & Carballedo, J. M. TH1 and TH2 lymphocyte development and regulation of TH cell-mediated immune responses of the skin. *J Invest Dermatol Symp Proc* **9**, 5-14 (2004).
- 22 Szabo, S. J. *et al.* A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655-669 (2000).
- 23 Lighvani, A. A. *et al.* T-bet is rapidly induced by interferon-gamma in lymphoid and myeloid cells. *Proc Natl Acad Sci U S A* **98**, 15137-15142 (2001).
- 24 Afkarian, M. *et al.* T-bet is a STAT1-induced regulator of IL-12R expression in naive CD4+ T cells. *Nat Immunol* **3**, 549-557, doi:10.1038/ni794 (2002).
- 25 Sher, A. & Coffman, R. L. Regulation of immunity to parasites by T cells and T cell-derived cytokines. *Annu Rev Immunol* **10**, 385-409, doi:10.1146/annurev.iy.10.040192.002125 (1992).

- 26 Zhang, D. H., Cohn, L., Ray, P., Bottomly, K. & Ray, A. Transcription factor GATA-3 is differentially expressed in murine Th1 and Th2 cells and controls Th2-specific expression of the interleukin-5 gene. *J Biol Chem* **272**, 21597-21603 (1997).
- 27 Abbas, A. K., Murphy, K. M. & Sher, A. Functional diversity of helper T lymphocytes. *Nature* **383**, 787-793 (1996).
- 28 Urban, J. F., Jr. *et al.* IL-13, IL-4, Ralpha, and Stat6 are required for the expulsion of the gastrointestinal nematode parasite *Nippostrongylus brasiliensis*. *Immunity* **8**, 255-264 (1998).
- 29 Parronchi, P. *et al.* IL-4 and IFN (alpha and gamma) exert opposite regulatory effects on the development of cytolytic potential by Th1 or Th2 human T cell clones. *Journal of immunology (Baltimore, Md : 1950)* **149**, 2977-2983 (1992).
- 30 Harrington, L. E. *et al.* Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* **6**, 1123-1132 (2005).
- 31 Mangan, P. R. *et al.* Transforming growth factor-beta induces development of the T(H)17 lineage. *Nature* **441**, 231-234 (2006).
- 32 Veldhoen, M., Hocking, R. J., Atkins, C. J., Locksley, R. M. & Stockinger, B. TGFbeta in the context of an inflammatory cytokine milieu supports de novo differentiation of IL-17-producing T cells. *Immunity* **24**, 179-189 (2006).
- 33 Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441**, 235-238 (2006).
- 34 Ivanov, I. I. *et al.* The orphan nuclear receptor RORgammat directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell* **126**, 1121-1133 (2006).
- 35 Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T(H)17 cells. *Nature* **448**, 484-487 (2007).
- 36 Yang, X. O. *et al.* T helper 17 lineage differentiation is programmed by orphan nuclear receptors ROR alpha and ROR gamma. *Immunity* **28**, 29-39 (2008).
- 37 Manel, N., Unutmaz, D. & Littman, D. R. The differentiation of human T(H)-17 cells requires transforming growth factor-beta and induction of the nuclear receptor RORgammat. *Nat Immunol* **9**, 641-649 (2008).
- 38 Infante-Duarte, C., Horton, H. F., Byrne, M. C. & Kamradt, T. Microbial lipopeptides induce the production of IL-17 in Th cells. *Journal of immunology (Baltimore, Md : 1950)* **165**, 6107-6115 (2000).
- 39 Ye, P. *et al.* Requirement of interleukin 17 receptor signaling for lung CXC chemokine and granulocyte colony-stimulating factor expression, neutrophil recruitment, and host defense. *J Exp Med* **194**, 519-527 (2001).
- 40 Huang, W., Na, L., Fidel, P. L. & Schwarzenberger, P. Requirement of interleukin-17A for systemic anti-*Candida albicans* host defense in mice. *J Infect Dis* **190**, 624-631 (2004).
- 41 Happel, K. I. *et al.* Divergent roles of IL-23 and IL-12 in host defense against *Klebsiella pneumoniae*. *J Exp Med* **202**, 761-769 (2005).
- 42 Dardalhon, V. *et al.* IL-4 inhibits TGF-beta-induced Foxp3+ T cells and, together with TGF-beta, generates IL-9+ IL-10+ Foxp3(-) effector T cells. *Nat Immunol* **9**, 1347-1355 (2008).
- 43 Veldhoen, M. *et al.* Transforming growth factor-beta 'reprograms' the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset. *Nat Immunol* **9**, 1341-1346 (2008).
- 44 Staudt, V. *et al.* Interferon-regulatory factor 4 is essential for the developmental program of T helper 9 cells. *Immunity* **33**, 192-202 (2010).
- 45 Anuradha, R. *et al.* IL-4-, TGF-beta-, and IL-1-dependent expansion of parasite antigen-specific Th9 cells is associated with clinical pathology in human lymphatic filariasis. *Journal of immunology (Baltimore, Md : 1950)* **191**, 2466-2473 (2013).
- 46 Duhon, T., Geiger, R., Jarrossay, D., Lanzavecchia, A. & Sallusto, F. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. *Nat Immunol* **10**, 857-863 (2009).
- 47 Trifari, S., Kaplan, C. D., Tran, E. H., Crellin, N. K. & Spits, H. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from T(H)-17, T(H)1 and T(H)2 cells. *Nat Immunol* **10**, 864-871 (2009).
- 48 Eyerich, S. *et al.* Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *J Clin Invest* **119**, 3573-3585 (2009).
- 49 Pennino, D. *et al.* IL-22 suppresses IFN-gamma-mediated lung inflammation in asthmatic patients. *J Allergy Clin Immunol* **131**, 562-570, doi:10.1016/j.jaci.2012.09.036 (2013).
- 50 Breitfeld, D. *et al.* Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *J Exp Med* **192**, 1545-1552 (2000).

- 51 Casamayor-Palleja, M., Khan, M. & MacLennan, I. C. A subset of CD4+ memory T cells contains preformed CD40 ligand that is rapidly but transiently expressed on their surface after activation through the T cell receptor complex. *J Exp Med* **181**, 1293-1301 (1995).
- 52 Chtanova, T. *et al.* T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. *Journal of immunology (Baltimore, Md : 1950)* **173**, 68-78 (2004).
- 53 Bryant, V. L. *et al.* Cytokine-mediated regulation of human B cell differentiation into Ig-secreting cells: predominant role of IL-21 produced by CXCR5+ T follicular helper cells. *Journal of immunology (Baltimore, Md : 1950)* **179**, 8180-8190 (2007).
- 54 Ozaki, K. *et al.* Regulation of B cell differentiation and plasma cell generation by IL-21, a novel inducer of Blimp-1 and Bcl-6. *Journal of immunology (Baltimore, Md : 1950)* **173**, 5361-5371 (2004).
- 55 Zotos, D. *et al.* IL-21 regulates germinal center B cell differentiation and proliferation through a B cell-intrinsic mechanism. *J Exp Med* **207**, 365-378 (2010).
- 56 Linterman, M. A. *et al.* IL-21 acts directly on B cells to regulate Bcl-6 expression and germinal center responses. *J Exp Med* **207**, 353-363 (2010).
- 57 Johnston, R. J. *et al.* Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science (New York, N Y)* **325**, 1006-1010 (2009).
- 58 Yu, D. *et al.* The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457-468 (2009).
- 59 Nurieva, R. I. *et al.* Bcl6 mediates the development of T follicular helper cells. *Science (New York, N Y)* **325**, 1001-1005 (2009).
- 60 Sakaguchi, S., Fukuma, K., Kuribayashi, K. & Masuda, T. Organ-specific autoimmune diseases induced in mice by elimination of T cell subset. I. Evidence for the active participation of T cells in natural self-tolerance; deficit of a T cell subset as a possible cause of autoimmune disease. *J Exp Med* **161**, 72-87 (1985).
- 61 Powrie, F. & Mason, D. OX-22high CD4+ T cells induce wasting disease with multiple organ pathology: prevention by the OX-22low subset. *J Exp Med* **172**, 1701-1708 (1990).
- 62 Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *Journal of immunology (Baltimore, Md : 1950)* **155**, 1151-1164 (1995).
- 63 Joetham, A. *et al.* Naturally occurring lung CD4(+)CD25(+) T cell regulation of airway allergic responses depends on IL-10 induction of TGF-beta. *Journal of immunology (Baltimore, Md : 1950)* **178**, 1433-1442 (2007).
- 64 Qureshi, O. S. *et al.* Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. *Science (New York, N Y)* **332**, 600-603 (2011).
- 65 Pandiyan, P., Zheng, L., Ishihara, S., Reed, J. & Lenardo, M. J. CD4+CD25+Foxp3+ regulatory T cells induce cytokine deprivation-mediated apoptosis of effector CD4+ T cells. *Nat Immunol* **8**, 1353-1362 (2007).
- 66 Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science (New York, N Y)* **299**, 1057-1061 (2003).
- 67 Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol* **4**, 330-336 (2003).
- 68 Chen, W. *et al.* Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. *J Exp Med* **198**, 1875-1886 (2003).
- 69 Horwitz, D. A., Zheng, S. G. & Gray, J. D. Natural and TGF-beta-induced Foxp3(+)CD4(+) CD25(+) regulatory T cells are not mirror images of each other. *Trends in immunology* **29**, 429-435 (2008).
- 70 Abbas, A. K. *et al.* Regulatory T cells: recommendations to simplify the nomenclature. *Nat Immunol* **14**, 307-308 (2013).
- 71 Apostolou, I. & von Boehmer, H. In vivo instruction of suppressor commitment in naive T cells. *J Exp Med* **199**, 1401-1408 (2004).
- 72 Kretschmer, K. *et al.* Inducing and expanding regulatory T cell populations by foreign antigen. *Nat Immunol* **6**, 1219-1227 (2005).
- 73 Chen, Y., Kuchroo, V. K., Inobe, J., Hafler, D. A. & Weiner, H. L. Regulatory T cell clones induced by oral tolerance: suppression of autoimmune encephalomyelitis. *Science (New York, N Y)* **265**, 1237-1240 (1994).

- 74 Groux, H. *et al.* A CD4<sup>+</sup> T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. *Nature* **389**, 737-742 (1997).
- 75 Miller, A., Lider, O., Roberts, A. B., Sporn, M. B. & Weiner, H. L. Suppressor T cells generated by oral tolerization to myelin basic protein suppress both in vitro and in vivo immune responses by the release of transforming growth factor beta after antigen-specific triggering. *Proc Natl Acad Sci U S A* **89**, 421-425 (1992).
- 76 Doetze, A. *et al.* Antigen-specific cellular hyporesponsiveness in a chronic human helminth infection is mediated by T(h)3/T(r)1-type cytokines IL-10 and transforming growth factor-beta but not by a T(h)1 to T(h)2 shift. *Int Immunol* **12**, 623-630 (2000).
- 77 Walker, M. R. *et al.* Induction of FoxP3 and acquisition of T regulatory activity by stimulated human CD4<sup>+</sup>CD25<sup>-</sup> T cells. *J Clin Invest* **112**, 1437-1443 (2003).
- 78 Walker, M. R., Carson, B. D., Nepom, G. T., Ziegler, S. F. & Buckner, J. H. De novo generation of antigen-specific CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells from human CD4<sup>+</sup>CD25<sup>-</sup> cells. *Proc Natl Acad Sci U S A* **102**, 4103-4108, doi:10.1073/pnas.0407691102 (2005).
- 79 Coombes, J. L. *et al.* A functionally specialized population of mucosal CD103<sup>+</sup> DCs induces Foxp3<sup>+</sup> regulatory T cells via a TGF-beta and retinoic acid-dependent mechanism. *J Exp Med* **204**, 1757-1764, doi:10.1084/jem.20070590 (2007).
- 80 Mucida, D. *et al.* Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. *Science (New York, N Y)* **317**, 256-260 (2007).
- 81 Sun, C.-M. *et al.* Small intestine lamina propria dendritic cells promote de novo generation of Foxp3<sup>+</sup> T reg cells via retinoic acid. *J Exp Med* **204**, 1775-1785 (2007).
- 82 Wang, J., Huizinga, T. W. & Toes, R. E. De novo generation and enhanced suppression of human CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells by retinoic acid. *J Immunol* **183**, 4119-4126, doi:10.4049/jimmunol.0901065 (2009).
- 83 Bluestone, J. A., Mackay, C. R., O'Shea, J. J. & Stockinger, B. The functional plasticity of T cell subsets. *Nature reviews. Immunology* **9**, 811-816, doi:10.1038/nri2654 (2009).
- 84 Mukasa, R. *et al.* Epigenetic instability of cytokine and transcription factor gene loci underlies plasticity of the T helper 17 cell lineage. *Immunity* **32**, 616-627, doi:10.1016/j.immuni.2010.04.016 (2010).
- 85 Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J Exp Med* **204**, 1849-1861, doi:10.1084/jem.20070663 (2007).
- 86 McGeachy, M. J. *et al.* TGF-beta and IL-6 drive the production of IL-17 and IL-10 by T cells and restrain T(H)-17 cell-mediated pathology. *Nat Immunol* **8**, 1390-1397, doi:10.1038/ni1539 (2007).
- 87 Acosta-Rodriguez, E. V., Napolitani, G., Lanzavecchia, A. & Sallusto, F. Interleukins 1beta and 6 but not transforming growth factor-beta are essential for the differentiation of interleukin 17-producing human T helper cells. *Nat Immunol* **8**, 942-949, doi:10.1038/ni1496 (2007).
- 88 Chen, Z., Tato, C. M., Muul, L., Laurence, A. & O'Shea, J. J. Distinct regulation of interleukin-17 in human T helper lymphocytes. *Arthritis and rheumatism* **56**, 2936-2946, doi:10.1002/art.22866 (2007).
- 89 Bending, D. *et al.* Highly purified Th17 cells from BDC2.5NOD mice convert into Th1-like cells in NOD/SCID recipient mice. *J Clin Invest* **119**, 565-572, doi:10.1172/JCI37865 (2009).
- 90 Lexberg, M. H. *et al.* Th memory for interleukin-17 expression is stable in vivo. *European journal of immunology* **38**, 2654-2664, doi:10.1002/eji.200838541 (2008).
- 91 Cosmi, L. *et al.* Identification of a novel subset of human circulating memory CD4(+) T cells that produce both IL-17A and IL-4. *J Allergy Clin Immunol* **125**, 222-230 e221-224, doi:10.1016/j.jaci.2009.10.012 (2010).
- 92 Deknuydt, F., Bioley, G., Valmori, D. & Ayyoub, M. IL-1beta and IL-2 convert human Treg into T(H)17 cells. *Clinical immunology* **131**, 298-307, doi:10.1016/j.clim.2008.12.008 (2009).
- 93 Komatsu, N. *et al.* Pathogenic conversion of Foxp3<sup>+</sup> T cells into TH17 cells in autoimmune arthritis. *Nat Med* **20**, 62-68, doi:10.1038/nm.3432 (2014).
- 94 Zhou, L. *et al.* TGF-beta-induced Foxp3 inhibits T(H)17 cell differentiation by antagonizing RORgamma function. *Nature* **453**, 236-240 (2008).
- 95 Zielinski, C. E. *et al.* Pathogen-induced human TH17 cells produce IFN-gamma or IL-10 and are regulated by IL-1beta. *Nature* **484**, 514-518 (2012).
- 96 Gagliani, N. *et al.* Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523**, 221-225 (2015).
- 97 Usui, T. *et al.* T-bet regulates Th1 responses through essential effects on GATA-3 function rather than on IFNG gene acetylation and transcription. *J Exp Med* **203**, 755-766 (2006).

- 98 Hegazy, A. N. *et al.* Interferons direct Th2 cell reprogramming to generate a stable GATA-3(+)T-bet(+) cell subset with combined Th2 and Th1 cell functions. *Immunity* **32**, 116-128 (2010).
- 99 Cope, A., Le Friec, G., Cardone, J. & Kemper, C. The Th1 life cycle: molecular control of IFN-gamma to IL-10 switching. *Trends in immunology* **32**, 278-286 (2011).
- 100 Glatman Zaretsky, A. *et al.* T follicular helper cells differentiate from Th2 cells in response to helminth antigens. *J Exp Med* **206**, 991-999 (2009).
- 101 King, I. L. & Mohrs, M. IL-4-producing CD4+ T cells in reactive lymph nodes during helminth infection are T follicular helper cells. *J Exp Med* **206**, 1001-1007 (2009).
- 102 Reinhardt, R. L., Liang, H.-E. & Locksley, R. M. Cytokine-secreting follicular T cells shape the antibody repertoire. *Nat Immunol* **10**, 385-393 (2009).
- 103 Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708-712, doi:10.1038/44385 (1999).
- 104 Clark, R. A. *et al.* The vast majority of CLA+ T cells are resident in normal skin. *Journal of immunology (Baltimore, Md : 1950)* **176**, 4431-4439 (2006).
- 105 Clark, R. A. *et al.* A novel method for the isolation of skin resident T cells from normal and diseased human skin. *J Invest Dermatol* **126**, 1059-1070 (2006).
- 106 Harrington, L. E., Janowski, K. M., Oliver, J. R., Zajac, A. J. & Weaver, C. T. Memory CD4 T cells emerge from effector T-cell progenitors. *Nature* **452**, 356-360 (2008).
- 107 Lohning, M. *et al.* Long-lived virus-reactive memory T cells generated from purified cytokine-secreting T helper type 1 and type 2 effectors. *J Exp Med* **205**, 53-61 (2008).
- 108 Chang, J. T. *et al.* Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science (New York, N Y)* **315**, 1687-1691 (2007).
- 109 Li, J., Huston, G. & Swain, S. L. IL-7 promotes the transition of CD4 effectors to persistent memory cells. *J Exp Med* **198**, 1807-1815 (2003).
- 110 Kondrack, R. M. *et al.* Interleukin 7 regulates the survival and generation of memory CD4 cells. *J Exp Med* **198**, 1797-1806 (2003).
- 111 Nagamine, K. *et al.* Positional cloning of the APECED gene. *Nat Genet* **17**, 393-398 (1997).
- 112 Finnish-German, A. C. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet* **17**, 399-403 (1997).
- 113 Romagnani, S. The role of lymphocytes in allergic disease. *J Allergy Clin Immunol* **105**, 399-408 (2000).
- 114 Turner, D. L. & Farber, D. L. Mucosal resident memory CD4 T cells in protection and immunopathology. *Frontiers in immunology* **5**, 331 (2014).
- 115 Pennock, N. D., Kedl, J. D. & Kedl, R. M. T Cell Vaccinology: Beyond the Reflection of Infectious Responses. *Trends in immunology* **37**, 170-180 (2016).
- 116 Vandenberg, L., Belmans, J., Van Woensel, M., Riva, M. & Van Gool, S. W. Exploiting the Immunogenic Potential of Cancer Cells for Improved Dendritic Cell Vaccines. *Frontiers in immunology* **6**, 663, doi:10.3389/fimmu.2015.00663 (2015).
- 117 Tan, A. C., Goubier, A. & Kohrt, H. E. A quantitative analysis of therapeutic cancer vaccines in phase 2 or phase 3 trial. *Journal for immunotherapy of cancer* **3**, 48, doi:10.1186/s40425-015-0093-x (2015).
- 118 Daniel, C., Weigmann, B., Bronson, R. & von Boehmer, H. Prevention of type 1 diabetes in mice by tolerogenic vaccination with a strong agonist insulin mimetope. *J Exp Med* **208**, 1501-1510, doi:10.1084/jem.20110574 (2011).
- 119 Wilson, D. R., Lima, M. T. & Durham, S. R. Sublingual immunotherapy for allergic rhinitis: systematic review and meta-analysis. *Allergy* **60**, 4-12, doi:10.1111/j.1398-9995.2005.00699.x (2005).
- 120 Radulovic, S., Wilson, D., Calderon, M. & Durham, S. Systematic reviews of sublingual immunotherapy (SLIT). *Allergy* **66**, 740-752, doi:10.1111/j.1398-9995.2011.02583.x (2011).
- 121 Nelson, H. S., Makatsori, M. & Calderon, M. A. Subcutaneous Immunotherapy and Sublingual Immunotherapy: Comparative Efficacy, Current and Potential Indications, and Warnings--United States Versus Europe. *Immunology and allergy clinics of North America* **36**, 13-24, doi:10.1016/j.iac.2015.08.005 (2016).
- 122 Creticos, P. S. *et al.* Nasal challenge with ragweed pollen in hay fever patients. Effect of immunotherapy. *J Clin Invest* **76**, 2247-2253, doi:10.1172/JCI112233 (1985).
- 123 Schleimer, R. P. *et al.* Regulation of human basophil mediator release by cytokines. I. Interaction with antiinflammatory steroids. *J Immunol* **143**, 1310-1317 (1989).
- 124 Walker, C., Virchow, J. C., Jr., Bruijnzeel, P. L. & Blaser, K. T cell subsets and their soluble products regulate eosinophilia in allergic and nonallergic asthma. *J Immunol* **146**, 1829-1835 (1991).

- 125 Chen, W. Y., Yu, J. & Wang, J. Y. Decreased production of endothelin-1 in asthmatic children after immunotherapy. *The Journal of asthma : official journal of the Association for the Care of Asthma* **32**, 29-35 (1995).
- 126 Meiler, F. *et al.* In vivo switch to IL-10-secreting T regulatory cells in high dose allergen exposure. *J Exp Med* **205**, 2887-2898, doi:10.1084/jem.20080193 (2008).
- 127 James, L. K. *et al.* Long-term tolerance after allergen immunotherapy is accompanied by selective persistence of blocking antibodies. *J Allergy Clin Immunol* **127**, 509-516 e501-505, doi:10.1016/j.jaci.2010.12.1080 (2011).
- 128 Novak, N. *et al.* Early suppression of basophil activation during allergen-specific immunotherapy by histamine receptor 2. *J Allergy Clin Immunol* **130**, 1153-1158 e1152, doi:10.1016/j.jaci.2012.04.039 (2012).
- 129 Uermosi, C. *et al.* IgG-mediated down-regulation of IgE bound to mast cells: a potential novel mechanism of allergen-specific desensitization. *Allergy* **69**, 338-347, doi:10.1111/all.12327 (2014).
- 130 Agache, I., Sugita, K., Morita, H., Akdis, M. & Akdis, C. A. The Complex Type 2 Endotype in Allergy and Asthma: From Laboratory to Bedside. *Current allergy and asthma reports* **15**, 29, doi:10.1007/s11882-015-0529-x (2015).
- 131 Mayumi, S. *et al.* Differential response in allergen-specific IgE, IgGs and IgA levels for predicting outcome of oral immunotherapy. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, doi:10.1111/pai.12535 (2016).
- 132 Lerner, A., Jeremias, P. & Matthias, T. The World Incidence and Prevalence of Autoimmune Diseases is Increasing. *International Journal of Celiac Disease* **3**, 151-155 (2015).
- 133 Wei, G. *et al.* Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* **35**, 299-311, doi:10.1016/j.immuni.2011.08.007 (2011).
- 134 Yosef, N. *et al.* Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496**, 461-468, doi:10.1038/nature11981 (2013).
- 135 Hawkins, R. D. *et al.* Global chromatin state analysis reveals lineage-specific enhancers during the initiation of human T helper 1 and T helper 2 cell polarization. *Immunity* **38**, 1271-1284, doi:10.1016/j.immuni.2013.05.011 (2013).
- 136 Meissner, F. & Mann, M. Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nat Immunol* **15**, 112-117, doi:10.1038/ni.2781 (2014).
- 137 Lonnberg, T., Chen, Z. & Laheesmaa, R. From a gene-centric to whole-proteome view of differentiation of T helper cell subsets. *Brief Funct Genomics* **12**, 471-482, doi:10.1093/bfpg/elt033 (2013).
- 138 Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587, doi:10.1038/nature13319 (2014).
- 139 Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575-581, doi:10.1038/nature13302 (2014).
- 140 Bini, L. *et al.* Extensive temporally regulated reorganization of the lipid raft proteome following T-cell antigen receptor triggering. *Biochem J* **369**, 301-309, doi:10.1042/BJ20020503 (2003).
- 141 Kobayashi, M., Katagiri, T., Kosako, H., Iida, N. & Hattori, S. Global analysis of dynamic changes in lipid raft proteins during T-cell activation. *Electrophoresis* **28**, 2035-2043, doi:10.1002/elps.200600675 (2007).
- 142 de Wet, B., Zech, T., Salek, M., Acuto, O. & Harder, T. Proteomic characterization of plasma membrane-proximal T cell activation responses. *J Biol Chem* **286**, 4072-4080, doi:10.1074/jbc.M110.165415 (2011).
- 143 Nyman, T. A. *et al.* A proteome database of human primary T helper cells. *Electrophoresis* **22**, 4375-4382, doi:10.1002/1522-2683(200112)22:20<4375::AID-ELPS4375>3.0.CO;2-P (2001).
- 144 Moulder, R. *et al.* Quantitative proteomics analysis of the nuclear fraction of human CD4+ cells in the early phases of IL-4-induced Th2 differentiation. *Mol Cell Proteomics* **9**, 1937-1953, doi:10.1074/mcp.M900483-MCP200 (2010).
- 145 Ruperez, P., Gago-Martinez, A., Burlingame, A. L. & Oses-Prieto, J. A. Quantitative phosphoproteomic analysis reveals a role for serine and threonine kinases in the cytoskeletal reorganization in early T cell receptor activation in human primary T cells. *Mol Cell Proteomics* **11**, 171-186, doi:10.1074/mcp.M112.017863 (2012).
- 146 Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386, doi:10.1038/nbt.1532 (2009).
- 147 Gundry, R. L. *et al.* A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol Cell Proteomics* **11**, 303-316, doi:10.1074/mcp.M112.018135 (2012).

- 148 Niehage, C. *et al.* The cell surface proteome of human mesenchymal stromal cells. *PLoS One* **6**, e20399, doi:10.1371/journal.pone.0020399 (2011).
- 149 Moest, H. *et al.* Malfunctioning of adipocytes in obesity is linked to quantitative surfaceome changes. *Biochim Biophys Acta* **1831**, 1208-1216 (2013).
- 150 Deeb, S. J., Cox, J., Schmidt-Supprian, M. & Mann, M. N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol Cell Proteomics* **13**, 240-251, doi:10.1074/mcp.M113.033977 (2014).
- 151 Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314, doi:10.1371/journal.pone.0121314 (2015).
- 152 Mirkowska, P. *et al.* Leukemia surfaceome analysis reveals new disease-associated features. *Blood* **121**, e149-159, doi:10.1182/blood-2012-11-468702 (2013).
- 153 Baek, S. *et al.* Therapeutic DC vaccination with IL-2 as a consolidation therapy for ovarian cancer patients: a phase I/II trial. *Cellular & molecular immunology* **12**, 87-95, doi:10.1038/cmi.2014.40 (2015).
- 154 McDermott, D. F. *et al.* The high-dose aldesleukin "select" trial: a trial to prospectively validate predictive models of response to treatment in patients with metastatic renal cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **21**, 561-568, doi:10.1158/1078-0432.CCR-14-1520 (2015).
- 155 Reichert, J. M. Antibodies to watch in 2016. *mAbs* **8**, 197-204, doi:10.1080/19420862.2015.1125583 (2016).
- 156 Zeng, Y., Ramya, T. N., Dirksen, A., Dawson, P. E. & Paulson, J. C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* **6**, 207-209, doi:10.1038/nmeth.1305 (2009).
- 157 Graessel, A. *et al.* A Combined Omics Approach to Generate the Surface Atlas of Human Naive CD4+ T Cells during Early T-Cell Receptor Activation. *Mol Cell Proteomics* **14**, 2085-2102, doi:10.1074/mcp.M114.045690 (2015).
- 158 Merl, J., Ueffing, M., Hauck, S. M. & von Toerne, C. Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Muller cells. *Proteomics* **12**, 1902-1911, doi:10.1002/pmic.201100549 (2012).
- 159 Vizcaino, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**, 223-226, doi:10.1038/nbt.2839 (2014).
- 160 Hauck, S. M. *et al.* Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol Cell Proteomics* **9**, 2292-2305, doi:10.1074/mcp.M110.001073 (2010).
- 161 Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res* **42**, D749-755, doi:10.1093/nar/gkt1196 (2014).
- 162 Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **42**, D191-198, doi:10.1093/nar/gkt1140 (2014).
- 163 Kumar, L. & M, E. F. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2**, 5-7 (2007).
- 164 Rigbolt, K. T., Vanselow, J. T. & Blagoev, B. GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Mol Cell Proteomics* **10**, O110 007450, doi:10.1074/mcp.O110.007450 (2011).
- 165 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
- 166 Boyle, E. I. *et al.* GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710-3715, doi:10.1093/bioinformatics/bth456 (2004).
- 167 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).
- 168 Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* **42**, D1091-1097, doi:10.1093/nar/gkt1068 (2014).
- 169 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210 (2002).
- 170 Pruitt, K. D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**, D756-763, doi:10.1093/nar/gkt1114 (2014).
- 171 Goldberg, T. *et al.* LocTree3 prediction of localization. *Nucleic Acids Res* **42**, W350-355, doi:10.1093/nar/gku396 (2014).

- 172 Kall, L., Krogh, A. & Sonnhammer, E. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21 Suppl 1**, i251-257, doi:10.1093/bioinformatics/bti1014 (2005).
- 173 R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (2015).
- 174 Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71-W74, doi:10.1093/nar/gkm306 (2007).
- 175 Barth, S. *et al.* Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res* **36**, 666-675, doi:10.1093/nar/gkm1080 (2008).
- 176 Bauer, C. *et al.* Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* **290**, 4801-4812, doi:10.1074/jbc.M114.605881 (2015).
- 177 Kohler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497 (1975).
- 178 Ausubel, F. M. *Current protocols in molecular biology*. (1989).
- 179 Dyballa, N. & Metzger, S. Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J Vis Exp* (2009).
- 180 Wefers, B. *et al.* Generation of targeted mouse mutants by embryo microinjection of TALEN mRNA. *Nat Protoc* **8**, 2355-2379, doi:10.1038/nprot.2013.142 (2013).
- 181 Brandl, C. *et al.* Creation of targeted genomic deletions using TALEN or CRISPR/Cas nuclease pairs in one-cell mouse embryos. *FEBS open bio* **5**, 26-35, doi:10.1016/j.fob.2014.11.009 (2015).
- 182 Nagy, A. *Manipulating the mouse embryo : a laboratory manual*. (2003).
- 183 Ittner, L. M. & Gotz, J. Pronuclear injection for the production of transgenic mice. *Nat Protoc* **2**, 1206-1215, doi:10.1038/nprot.2007.145 (2007).
- 184 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).
- 185 Monks, C. R., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82-86 (1998).
- 186 Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science (New York, N Y)* **285**, 221-227 (1999).
- 187 Freiberg, B. A. *et al.* Staging and resetting T cell activation in SMACs. *Nat Immunol* **3**, 911-917 (2002).
- 188 Yokosuka, T. *et al.* Spatiotemporal regulation of T cell costimulation by TCR-CD28 microclusters and protein kinase C theta translocation. *Immunity* **29**, 589-601 (2008).
- 189 Hegde, V. L., Singh, N. P., Nagarkatti, P. S. & Nagarkatti, M. CD44 mobilization in allogeneic dendritic cell-T cell immunological synapse plays a key role in T cell activation. *J Leukoc Biol* **84**, 134-142, doi:10.1189/jlb.1107752 (2008).
- 190 Naramura, M. *et al.* c-Cbl and Cbl-b regulate T cell responsiveness by promoting ligand-induced TCR down-modulation. *Nat Immunol* **3**, 1192-1199, doi:10.1038/ni855 (2002).
- 191 Gowans, J. L. & Knight, E. J. THE ROUTE OF RE-CIRCULATION OF LYMPHOCYTES IN THE RAT. *Proc R Soc Lond B Biol Sci* **159**, 257-282 (1964).
- 192 Mempel, T. R., Henrickson, S. E. & Von Andrian, U. H. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature* **427**, 154-159 (2004).
- 193 Friedman, R. S., Beemiller, P., Sorensen, C. M., Jacobelli, J. & Krummel, M. F. Real-time analysis of T cell receptors in naive cells in vitro and in vivo reveals flexibility in synapse and signaling dynamics. *J Exp Med* **207**, 2733-2749 (2010).
- 194 Mitchell, C. J. *et al.* A multi-omic analysis of human naive CD4+ T cells. *BMC systems biology* **9**, 75, doi:10.1186/s12918-015-0225-4 (2015).
- 195 Zhang, L. *et al.* Proteomic analysis of mouse liver plasma membrane: use of differential extraction to enrich hydrophobic membrane proteins. *Proteomics* **5**, 4510-4524 (2005).
- 196 Josic, D. *et al.* Use of selective extraction and fast chromatographic separation combined with electrophoretic methods for mapping of membrane proteins. *Electrophoresis* **26**, 2809-2822 (2005).
- 197 Michie, C. A., McLean, A., Alcock, C. & Beverley, P. C. Lifespan of human lymphocyte subsets defined by CD45 isoforms. *Nature* **360**, 264-265 (1992).
- 198 McFarland, R. D., Douek, D. C., Koup, R. A. & Picker, L. J. Identification of a human recent thymic emigrant phenotype. *Proc Natl Acad Sci U S A* **97**, 4215-4220, doi:10.1073/pnas.070061597 (2000).

- 199 Risso, A. *et al.* CD69 in resting and activated T lymphocytes. Its association with a GTP binding protein and biochemical requirements for its expression. *Journal of immunology (Baltimore, Md : 1950)* **146**, 4105-4114 (1991).
- 200 da Cunha, J. P. *et al.* Bioinformatics construction of the human cell surfaceome. *Proc Natl Acad Sci U S A* **106**, 16752-16757, doi:10.1073/pnas.0907939106 (2009).
- 201 Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141-1149, doi:10.1002/pmic.200900258 (2010).
- 202 Trompette, A. *et al.* Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med* **20**, 159-166, doi:10.1038/nm.3444 (2014).
- 203 Cebrian, M. *et al.* Triggering of T cell proliferation through AIM, an activation inducer molecule expressed on activated human lymphocytes. *J Exp Med* **168**, 1621-1637 (1988).
- 204 Suzuki, K. *et al.* Semaphorin 7A initiates T-cell-mediated inflammatory responses through alpha1beta1 integrin. *Nature* **446**, 680-684 (2007).
- 205 Sinclair, L. V. *et al.* Control of amino-acid transport by antigen receptors coordinates the metabolic reprogramming essential for T cell differentiation. *Nat Immunol* **14**, 500-508, doi:10.1038/ni.2556 (2013).
- 206 Hayashi, K., Jutabha, P., Endou, H., Sagara, H. & Anzai, N. LAT1 is a critical transporter of essential amino acids for immune reactions in activated human T cells. *J Immunol* **191**, 4080-4085, doi:10.4049/jimmunol.1300923 (2013).
- 207 Hayashi, K., Ouchi, M., Endou, H. & Anzai, N. HOXB9 acts as a negative regulator of activated human T cells in response to amino acid deficiency. *Immunol Cell Biol*, doi:10.1038/icb.2016.13 (2016).
- 208 Kayes, L. M. *et al.* Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients. *Am J Hum Genet* **54**, 424-436 (1994).
- 209 Jenne, D. E. *et al.* A common set of at least 11 functional genes is lost in the majority of NF1 patients with gross deletions. *Genomics* **66**, 93-97, doi:10.1006/geno.2000.6179 (2000).
- 210 Buchberg, A. M., Bedigian, H. G., Jenkins, N. A. & Copeland, N. G. Evi-2, a common integration site involved in murine myeloid leukemogenesis. *Mol Cell Biol* **10**, 4658-4666 (1990).
- 211 Matesanz-Isabel, J. *et al.* New B-cell CD molecules. *Immunol Lett* **134**, 104-112, doi:10.1016/j.imlet.2010.09.019 (2011).
- 212 Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730 (1999).
- 213 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
- 214 Rugg-Gunn, P. J. *et al.* Cell-surface proteomics identifies lineage-specific markers of embryo-derived stem cells. *Dev Cell* **22**, 887-901, doi:10.1016/j.devcel.2012.01.005 (2012).
- 215 Boheler, K. R. *et al.* A human pluripotent stem cell surface N-glycoproteome resource reveals markers, extracellular epitopes, and drug targets. *Stem Cell Reports* **3**, 185-203, doi:10.1016/j.stemcr.2014.05.002 (2014).
- 216 Schiller, H. B. *et al.* Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Molecular systems biology* **11**, 819, doi:10.15252/msb.20156123 (2015).
- 217 Strassberger, V. *et al.* A comprehensive surface proteome analysis of myeloid leukemia cell lines for therapeutic antibody development. *J Proteomics* **99**, 138-151, doi:10.1016/j.jprot.2014.01.022 (2014).
- 218 Ferreras, C., Lancrin, C., Lie, A. L. M., Kouskoff, V. & Lacaud, G. Identification and characterization of a novel transcriptional target of RUNX1/AML1 at the onset of hematopoietic development. *Blood* **118**, 594-597, doi:10.1182/blood-2010-06-294124 (2011).
- 219 Halim, A., Nilsson, J., Ruetschi, U., Hesse, C. & Larson, G. Human urinary glycoproteomics; attachment site specific analysis of N- and O-linked glycosylations by CID and ECD. *Mol Cell Proteomics* **11**, M111 013649, doi:10.1074/mcp.M111.013649 (2012).
- 220 Kim, S. Y. *et al.* A novel antibody-drug conjugate targeting SAIL for the treatment of hematologic malignancies. *Blood cancer journal* **5**, e316, doi:10.1038/bcj.2015.39 (2015).
- 221 Forconi, F. & Moss, P. Perturbation of the normal immune system in patients with CLL. *Blood* **126**, 573-581 (2015).
- 222 Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425-440, doi:10.1016/j.cell.2015.06.043 (2015).
- 223 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539, doi:10.1093/nar/gkj109 (2006).

- 
- 224 Kawano, H. *et al.* A novel feedback mechanism by Ephrin-B1/B2 in T-cell activation involves a concentration-dependent switch from costimulation to inhibition. *European journal of immunology* **42**, 1562-1572, doi:10.1002/eji.201142175 (2012).
- 225 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

## Materials

### 1. Reagents, media, buffers, enzymes, cell lines and kits

Reagent	Company
1-step Ultra TMB-ELISA substrate	Thermo Fisher Scientific, Darmstadt, Germany
5-Brom-4-chlor-3-indoylphosphat	Sigma-Aldrich, Taufkirchen, Germany
Accu Prime Pfx DNA Polymerase	Invitrogen by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Acetone	Merck, Darmstadt, Germany
Acetonitrile	Sigma-Aldrich, Taufkirchen, Germany
Acrylamid (Rotiphorese Gel 30)	Carl Roth, Karlsruhe, Germany
Adjuvant CPG2006	TIB MOLBIOL, Berlin, Germany
Agar	AppliChem GmbH, Darmstadt, Germany
Agarose DNA	Biozym Scientific GmbH, Hessisch Oldendorf, Germany
AIMV serum-free medium	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Albumin from bovine serum (BSA)	Sigma-Aldrich, Taufkirchen, Germany
aluminium-sulfate-(14-18)-hydrate	Sigma-Aldrich, Taufkirchen, Germany
Ambion Mega Short Script Kit	Ambion, Thermo Fisher Scientific, Darmstadt, Germany
Ambion Megaclear Kit	Ambion, Thermo Fisher Scientific, Darmstadt, Germany
Ambion mMESSAGE mMACHINE_T7 Ultra Transcription Kit	Ambion, Thermo Fisher Scientific, Darmstadt, Germany
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare, Freiburg, Germany
Aminoxy-Biotin	Biotium Inc., Hayward, CA, USA
Aminopterin	Sigma-Aldrich, Taufkirchen, Germany
Ammoniumpersulfate	Sigma-Aldrich, Taufkirchen, Germany
Ampicillin	Carl Roth, Karlsruhe, Germany
Aniline	Sigma-Aldrich, Taufkirchen, Germany
ArC Amine reactive beads	BD Biosciences, Heidelberg, Germany
Ascl	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
AsiSI	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
autoMACS rinsing solution	Miltenyi Biotech, Bergisch-Gladbach, Germany
autoMACS running buffer	Miltenyi Biotech, Bergisch-Gladbach, Germany
BbsI	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
BD Comp beads (anti-mouse Ig kappa)	BD Biosciences, Heidelberg, Germany
BD Comp beads (anti-rat Ig kappa)	BD Biosciences, Heidelberg, Germany
BD Cytifix/Cytoperm Kit	BD Biosciences, Heidelberg, Germany

Bio Non-fat dry milk powder	Heirler Cenovis GmbH, Radolfzell, Germany
CD4+ CD62L+ T cell Isolation Kit II, mouse	Miltenyi Biotech, Bergisch-Gladbach, Germany
CD45RO microbeads	Miltenyi Biotech, Bergisch-Gladbach, Germany
Cellfectine II Reagent	Thermo Fisher Scientific, Darmstadt, Germany
Chemo-competent NEB5alpha	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
Coomassie Brilliant Blue - G250	Sigma-Aldrich, Taufkirchen, Germany
Coomassie-Brilliant-Blue R-250	Thermo Fisher Scientific, Darmstadt, Germany
Custom-made c16orf54 peptides for immunization	Peps4LS, Heidelberg, Germany
D-PBS	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
DEPC-treated water (pyrogen-free)	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Dimethylformamide	Sigma-Aldrich, Taufkirchen, Germany
Dimethylsulfoxid (DMSO), cell culture grad	Applichem, Darmstadt, Germany
Dithiothreitol (DTT)	Thermo Fisher Scientific, Darmstadt, Germany
DMEM	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
DNA Ligase Buffer (10 x)	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
DNA Loading Dye (6 x)	Thermo Fisher Scientific, Darmstadt, Germany
Dpnl	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
DreamTaq DNA Polymerase, dNTP-Mix	Thermo Fisher Scientific, Darmstadt, Germany
EconoTaq PLUS Master Mix (2 x)	Lucigen, Middleton, WI, USA
EDTA (0.05 %, pH 8)	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
EDTA disodium salt dihydrate	Sigma-Aldrich, Taufkirchen, Germany
EDTA-free complete protease inhibitor cocktail	Roche Diagnostics, Mannheim, Germany
Embryo-tested water	Sigma-Aldrich, Taufkirchen, Germany
Ethanol absolute	Merck, Darmstadt, Germany
Ethidium bromide solution	Sigma-Aldrich, Taufkirchen, Germany
Fast Digest Nhe I	Thermo Fisher Scientific, Darmstadt, Germany
Fast Digest Not I	Thermo Fisher Scientific, Darmstadt, Germany
Fast Digest Xba I	Thermo Fisher Scientific, Darmstadt, Germany
Fast Digest Xho I	Thermo Fisher Scientific, Darmstadt, Germany
Fast Digest Green Buffer (10 x)	Thermo Fisher Scientific, Darmstadt, Germany
FastStart Universal SYBR Green Mastermix 2 x (ROX)	Roche Diagnostics, Mannheim, Germany
Fetal calf serum (FCS) Hyclone II	Perbio Science, Bonn, Germany
Formic acid	Sigma-Aldrich, Taufkirchen, Germany
G418 sulfate 50 mg/mL	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
G7 buffer	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
Gene Expression Hybridization Kit	Agilent Technologies, Santa Clara, USA
GeneJet gel extraction Kit	Thermo Fisher Scientific, Darmstadt, Germany
GeneJet Plasmid Miniprep Kit	Thermo Fisher Scientific, Darmstadt, Germany
Gentamicin	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Glycerol	Sigma-Aldrich, Taufkirchen, Germany
Glycine	Sigma-Aldrich, Taufkirchen, Germany

HEK-293 cell line	ATCC, Manassas, VA, USA
Heparin- Natrium 250.000 U	Ratiopharm, Ulm Germany
Herculase II Fusion DNA Polymerase	Agilent Technologies, Santa Clara, CA, USA)
Herculase II reaction buffer (5 x)	Agilent Technologies, Santa Clara, CA, USA)
High capacity cDNA reverse transcription Kit	Applied Biosystems, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
HT (hypoxantin-aminopterin) supplement	Life Technologies, Thermo Fisher Scientific, Darmstadt, Germany
Hybridoma cloning factor	Capricorn Scientific, Ebsdorfergrund, Germany
Hydrochloric acid (HCl)	Merck, Darmstadt, Germany
Imidazole	Sigma-Aldrich, Taufkirchen, Germany
Insect Xpress protein-free insect cell medium	Lonza, Basel, Switzerland
Iodacetamide	Sigma-Aldrich, Taufkirchen, Germany
Isopropanol	Carl Roth, Karlsruhe, Germany
KSOM medium	Merck Millipore, Darmstadt, Germany
L-Glutamine	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
LEGENDScreen Human Cell Screening (PE) Kit	BioLegend, San Diego, CA, USA
Lightning-Link PE-TexasRed Tandem Conjugation Kit	Innova Biosciences, Cambridge, UK
Lipofectamine 3000 Transfection Reagent, P3000 Reagent	Thermo Fisher Scientific, Darmstadt, Germany
LIVE/DEAD® Fixable Aqua Dead Cell Stain Kit, for 405 nm excitation	Molecular probes by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Low Input Quick Amp Labeling Kit	Agilent Technologies, Santa Clara, USA
Lymphoprep	Fresenius Kabi Norge AS for Axis-Shield PoC AS, Oslo, Norway
MagStrep type 2HC beads and corresponding buffer system (Washing buffer W, Elution Buffer BE)	Iba lifesciences, Göttingen, Germany
Methanol	Sigma-Aldrich, Taufkirchen, Germany
MOPS SDS running buffer	NuPAGE Novex, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Neuro-2A cell line (ATCC CCL-131)	Kind gift from AG Krappman, Toxicology, Helmholtz Center Munich
Naive CD4+ T cell isolation Kit II human	Miltenyi Biotech, Bergisch-Gladbach, Germany
NEB buffer 2 and 2.1	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
NEB CutSmart Buffer	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
Nitrotetrazolium Blue chloride	Sigma-Aldrich, Taufkirchen, Germany
Non-essential amino acids (NEAA)	PAA, Linz, Austria
Non-fat dry milk powder	Frema Reform, granoVita, Heimertingen, Germany
NP-40	Sigma-Aldrich, Taufkirchen, Germany
Nucleo Spin Plasmid Kit	Marchery Nagel, Düren, Germany
NuPAGE LDS Sample Buffer	Novex, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
NuPage Transfer buffer (10 x)	NuPAGE Novex, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Opti-MEM	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
orthophosphoric acid (85 %)	Sigma-Aldrich, Taufkirchen, Germany
pAcGP67-B vector	BD Biosciences, Heidelberg, Germany

PBS with CaCl <sub>2</sub> / MgCl <sub>2</sub>	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
pbs-U6_chimaericRNA vector	provided by the IDG (Helmholtz Center Munich), also used for WB
pCAG-Cas9-162pA	provided by the IDG (Helmholtz Center Munich), also used for WB
pCas9-T2A-GFP vector	provided by the IDG (Helmholtz Center Munich), also used for WB
pcDNA3.1 vector	Invitrogen, Thermo Fisher Scientific, Darmstadt, Germany
PEG 1500	Roche Diagnostics, Mannheim, Germany
Penicillin-Streptomycin	PAA, Linz, Austria
Plasmid Maxi Kit	Qiagen, Hilden, Germany
PNGase F	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
ProGreen Baculovirus DNA	Ab Vector, San Diego, CA, USA
Propidium iodide	Sigma-Aldrich, Taufkirchen, Germany
Proteinase K	Sigma-Aldrich, Taufkirchen, Germany
Qiagen gel extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR purification Kit	Qiagen, Hilden, Germany
rh IL-2 Proleukin	Novartis, Basel, Switzerland
RNA 6000 Nano Kit	Agilent Technologies, Santa Clara, USA
RNeasy Mini Kit for RNA Isolation	Qiagen, Hilden, Germany
RPMI 1640	PAA, Linz, Austria
RPMI 1640	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
SCR7	Xcess Biosciences Inc, San Diego, CA, USA
SDS solution (10 %)	gibco by LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Seeblue prestained protein standard	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Sequencing-grade modified trypsin	Promega, Madison, WI, USA
Sf9 insect cells	Thermo Fisher Scientific, Darmstadt, Germany
SOC-medium	Invitrogen, Thermo Fisher Scientific, Darmstadt, Germany
Sodium azide (NaN <sub>3</sub> )	Merck, Darmstadt, Germany
Sodium carbonate (Na <sub>2</sub> CO <sub>3</sub> )	Merck, Darmstadt, Germany
Sodium chloride (NaCl)	Carl Roth, Karlsruhe, Germany
Sodium dodecyl sulfate	Sigma-Aldrich, Taufkirchen, Germany
Sodium hydroxide (NaOH)	Merck, Darmstadt, Germany
Sodium periodate (NaIO <sub>4</sub> )	Sigma-Aldrich, Taufkirchen, Germany
Sodium-Pyruvate (C <sub>3</sub> H <sub>3</sub> NaO <sub>3</sub> )	PAA, Linz, Austria
Strep-Tactin Superflow 50 % suspension	Iba lifesciences, Göttingen, Germany
Sulphuric acid (H <sub>2</sub> SO <sub>4</sub> )	Merck, Darmstadt, Germany
SurePrint G3 Gene Expression 8x60K microarray	Agilent Technologies, Santa Clara, USA
T cell activation/expansion Kit	Miltenyi Biotech, Bergisch-Gladbach, Germany
T4 DNA Ligase	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
T7-endonuclease	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
TEMED	Carl Roth, Karlsruhe, Germany

Trichloroacetic acid	Merck, Darmstadt, Germany
Tricin	Carl Roth, Karlsruhe, Germany
Tris(hydroxymethyl)-aminomethan	Sigma-Aldrich, Taufkirchen, Germany
Triton X	Sigma-Aldrich, Taufkirchen, Germany
Trizma base (Sigma ultra)	Sigma-Aldrich, Taufkirchen, Germany
Trizma hydrochloride (Sigma ultra)	Sigma-Aldrich, Taufkirchen, Germany
Trypanblue 0.4 % solution	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Trypsin 0.05 % EDTA	Sigma-Aldrich, Taufkirchen, Germany
Tryptone	AppliChem GmbH, Darmstadt, Germany
Tween 20 detergent	Calbiochem, San Diego, CA, USA
Urea	Sigma-Aldrich, Taufkirchen, Germany
Wizard Genomic DNA purification Kit	Promega, Madison, WI, USA
Wizard Genomic DNA purification Kit	Promega, Madison, WI, USA
XbaI	NEB-New England Biolabs GmbH, Frankfurt a.M., Germany
XtremeGene HP DNA Transfection Reagent	Roche Diagnostics, Mannheim, Germany
Yeast extract	AppliChem GmbH, Darmstadt, Germany

**Table M 1:** List of reagents, media, buffer, enzymes and kits.

## 2. Antibodies

Antibody	Dilution	Company
anti-CD3 NA/LE purified (clone UCHT1)	0.75 µg/mL	BD Biosciences, Heidelberg, Germany
anti-CD28 NA/LE purified (clone 28.2)	0.75 µg/mL	BD Biosciences, Heidelberg, Germany
anti-CD3-PerCPCy5.5	1:100	BD Biosciences, Heidelberg, Germany
anti-CD4-APCCy7	1:20	BD Biosciences, Heidelberg, Germany
anti-CD8-BV711	1:200	BioLegend, San Diego, CA, USA
anti-CD11a-FITC	1:100	BD Biosciences, Heidelberg, Germany
anti-CD14-AlexaFlour700	1:100	BD Biosciences, Heidelberg, Germany
anti-CD19-BV605	1:20	BD Biosciences, Heidelberg, Germany
anti-CD25-PECy7	1:100	BD Biosciences, Heidelberg, Germany
anti-CD45RA-BV450	1:50	BD Biosciences, Heidelberg, Germany
anti-CD45RO-BV650	1:100	BioLegend, San Diego, CA, USA
anti-CD56-PECy7	1:100	BioLegend, San Diego, CA, USA
anti-CD62L-PE-TexasRed	1:100	BD Biosciences, Heidelberg, Germany
anti-CD69-PE	1:100	BD Biosciences, Heidelberg, Germany
mouse-anti-rat IgG2A (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB
mouse-anti-rat IgG2C (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB
mouse-anti-rat IgG1 (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB

rat-anti-mouse IgG (H+L)-AlexaFlour594	1:200	Dianova, Hamburg, Germany
rat-anti-mouse IgG2a (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB
rat-anti-mouse IgG2b (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB
rat-anti-mouse IgG1 (hybridoma SN)	1:10	provided by the IMI (Helmholtz Center Munich), also used for WB
mouse-anti-rat IgG (H+L) Affini-Pure -DyLight594	1:200	Jackson ImmunoResearch, West Grove, PA, USA
Streptavidin-PE	1:100	BD Biosciences, Heidelberg, Germany

**Table M 2:** List of antibodies used for T cell stimulation or flow cytometry application.

1 <sup>st</sup> Antibody	Dilution	Company
anti-EVI2a	1:300	proteintech, Rosemont, IL, USA
anti-RNF149	1:150	Sigma-Aldrich, Taufkirchen, Germany
anti-TMEM2	1:1000	Abcam, Cambridge, UK
anti-NPTN	1:500	Novus biological, Littleton, CO, USA
anti-beta-Actin	1:10000	santa cruz biotechnology, Heidelberg, Germany
V5 epitope tag antibody	1:5000	Invitrogen, Thermo Fisher Scientific, Darmstadt, Germany
Strep-Tactin AP conjugate	1:4000	iba lifesciences, Göttingen, Germany

**Table M 3:** List of primary antibodies used for Western blot application.

2 <sup>nd</sup> Antibody	Dilution	Company
goat-a-rabbit	1:3000	santa cruz biotechnology, Heidelberg, Germany
goat-a-mouse	1:2000-1:5000	santa cruz biotechnology, Heidelberg, Germany
mouse-anti-rat IgG2A-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)
mouse-anti-rat IgG2C-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)
mouse-anti-rat IgG1-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)
rat-anti-mouse IgG2a-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)
rat-anti-mouse IgG2b-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)
rat-anti-mouse IgG1-HRP	1:1000	provided by the IMI (Helmholtz Center Munich)

**Table M 4:** List of secondary antibodies used for Western blot application.

### 3. Primer and oligos

Primer/oligo	Sequence (5' - 3') (ordered from Metabion International AG, Munich, Germany)
qPCR:	
APOC3 fw	TTTCAGGGAAGCTGAAGCCAT
APOC3 rev	CTCCCTTCTCAGCTTCATGC

ATP1A1 fw	CCCTTTTGGCCTTTTATCA
ATP1A1 rev	GGAGCTGCTCTGTGCTTTTC
c16orf54 fw	CTCCAAGCCTGCACCTCTTA
c16orf54 rev	CGGCATCTTGGGTGAGTCC
CALML5 fw	GAGTTTCCTTAGCTGGGCCT
CALML5 rev	TACAAAAGGCTTTCTCCGC
CALML5 fw1	CAAGGTCTGGAGGCAGAGAG
CALML5 rev1	CTGCTTTTGTGCTGGGACTC
DCD fw	TAACCCTGGGTCTTACCTG
DCD rev	CCCTGGTCTGTGCCTATGAT
DCD fw1	CCTGCATTTTCCTTTTGAGC
DCD rev1	CCTCTTCCTGACAGCTCTGG
ECE fw	CCGAGCCTCTCAATCAACTC
ECE rev	AATCATCAAGCACCTCCTCG
EF-1alpha fw	CTGAACCATCCAGGCCAAAT
EF-1alpha rev	GCCCTGTGGCAATCCAAT
ERO1L fw	TGTAGTCTTGGGAAAAGCCTG
ERO1L rev	GAGACAGCGGCACAGAGGT
EVI2A fw	GTTTGCTTTTGTTCAGGAGA
EVI2A rev	CAGATTTTGACCAAGCATTTTG
GOT2 fw	CTTCAGTGACTCCAGAATGG
GOT2 rev	ATGGCCCTGCTGCACTC
GPA33 fw	GGAGTTTCCACAGAGATGGC
GPA33 rev	GGCAGAGAAGAAGCAAGACC
INHBB fw	TTTCAGGTAAAGCCACAGGC
INHBB rev	GCGTTTCCGAAATCATCAG
NPTN fw	CACTCTGGCCTTTCTTGGGA
NPTN rev	CATTGGTCCAGCGTCAGGAA
RNF149 fw	GAATTCCTTTTGCACCAGC
RNF149 rev	GCTACGGGAACATCACCTTG
SBSN fw	TGGACAGGGCAACCATCAAA
SBSN rev	ATGAAAGGCGTGTGACCGA
SLC1A4 fw	GCTGTTCTGGGTACGACTT
SLC1A4 rev	CCCAAAGAGACGGTGGACT
SLC25A3 fw	CCACGAACACCATCCTTTT
SLC25A3 rev	GGTGGGTCTTAAGTTGTGG
SLC25A5 fw	TCTGCAGTGATCTGCTTGCT
SLC25A5 rev	TTCAACATGACAGATGCCG
SLC4A7 fw	CAGTTCATCCATTTCCGTGA
SLC4A7 rev	AAGATGGACGGGAATCTCCT
SLC5A3 fw	AAGGTGGTGGTTCGAATCTG
SLC5A3 rev	CCACAGGATTGTTTTGGGTC
SLC5A6 fw	CTTCCAGCCACAGTCTCACA

SLC5A6 rev	TCAGAAAAGGGAGCGATGTT
SLC7A1 fw	AGAGGACAGCCTCGATCTTG
SLC7A1 rev	GTCTGTCTGTTTCGCGATCCT
SMR3B fw	TTTGCCGTCTTCAACTGGC
SMR3B rev	AGTTGCCTCTTTGTGTGAGGA
SYPL1 fw	ATGGCGCCCAACATCTACTT
SYPL1 rev	AAGCAATCCACTCGAGGACC
TMEM2 fw	ATAGCCAGATGGGTGACGAC
TMEM2 rev	CACCTGAGTTAACGGACGCT
<b>cloning:</b>	
c16orf54_P1	GATCGCTAGCATGCCGTTGACTCCAGAGCCG
c16orf54_P2	GATCCTCGAGTCAGAACCCACACTGGTCCG
c16orf54_P3	GATCGCTAGCATGAGCGCTGAGCCACCCATT CGAAAAAGTGACTTCATGCCCTGTGGGC
c16orf54_P4	GATCCTCGATCATTTTTTGAAGTGCGGGTGGC TCCAGCGCTGAACCCACATGGTCCGGCCTTC
c16orf54_P5	GATCTCTAGAGGGATCGAGGGAAGGGAACGC CTGTTCCGCCGTG
c16orf54_P6	GATCGCGCCGCTCAGAACCCACACTGGTCCG
c16orf54_P7	GATCCCCGGGAGCGCTTGGAGCCACCCGCAGT TCGAAAAAGAAGCCTGTTCCGCCGTG
T7_seq_fw	TAATACGACTCACTATAGGG (standard sequencing primer of GATC Biotech)
pAcGP67- B_seq_fw	CTAGTAAATCAGTCACACCAAGG
<b>cloning (CRISPR/CAS):</b>	
primer T7 FW (IDG)	GTACAAAATACGTGACGTAGAAAAG
primer Tracr RV (IDG)	AAAAAAAGCACCGACTCGGTG
Primer_ T7_Tracr-1	GTACAAAATACGTGACGTAGAAAAG
Primer_ T7_Tracr-2	AAAAAAAGCACCGACTCGGTG
primer U6 (IDG)	GAGGGCCTATTTCCCATG
mCRISPR 1_chimA	CACCGGAGCCCCTCCGTGTGCCAGA
mCRISPR 1_chimB	AAACTCTGGGCACACGGAGGGGCTCC
mCRISPR 3_chimA	CACCGCGGGCGGCGGAACAGGCGTT
mCRISPR 3_chimB	AAACAACGCCTGTTCCGCCGCCGCC
mT7_CRISPR 1_chimA	CACCTAATACGACTCACTATAGGGAGCCCCCTCCGTGTGCCAGA
mT7_CRISPR 1_chimB	AAACTCTGGGCACACGGAGGGGCTCCCTATAGTGAGTCGTATTA
mT7_CRISPR 3_chimA	CACCTAATACGACTCACTATAGGGCGGGCGGCGGAACAGGCGTT
mT7_CRISPR 3_chimB	AAACAACGCCTGTTCCGCCGCCGCCCTATAGTGAGTCGTATTA

long_Primer_m1_fw	TAATACGACTCACTATAGGAGCCCCTCCGTGTGCCAGAGTTTAAGAGCTATGCTGGAAACAGC
<b>Genotyping:</b>	
CRISPR_checkP1_fw	ATGATGTGCAGTCCCAGTGA
CRISPR_checkP1_rev	GGCCTTCCTCTTCCAGAAT
CRISPR_checkP2_fw	AGCCCAGACCTTTGATCTGT
CRISPR_checkP2_rev	GAGCCGTACCAATCCTCAGA
CRISPR_checkP3_fw	ACTGTCTCATCCTGTGCCAC
CRISPR_checkP3_rev	TGGGCTCAATCCACAGTTCTC
LacZ_fw	AATGATGATTTAGCCGCGC
LacZ rev	CCGCCAAGACTGTTACCCAT
AI467606_1fw	ACGGCGAAATCCAAGATTGA
AI467606_1rev	TGGGAGAAGGAAAGGACAGC

**Table M 5:** List of primer names and corresponding sequences.

#### 4. Buffers and stocks (self-made)

Agent	Concentration / amount
<b>Ampicillin stock solution</b>	
20 mg/ml Ampicillin	1 g
70 % Ethanol	35 mL
ddH <sub>2</sub> O	15 mL
<b>AP-detection buffer</b>	
100 mM Tris	12.11 g
10 mM MgCl <sub>2</sub> *6 H <sub>2</sub> O	2.033 g
100 mM NaCl	5.845 g
ad ddH <sub>2</sub> O, pH 9.5	1 L
<b>AP-substrate solution</b>	
10 x NBT stock solution	125 µL
BCIP stock solution	125 µL
ad AP-detection buffer	12.5 mL
<b>APS stock solution</b>	
10 % Ammoniumpersulfate	0.1 g
ad ddH <sub>2</sub> O	1 mL
<b>BCIP stock solution</b>	
0.5 % 5-Brom-4-chlor-3-indoylphosphat	0.1 g
Dimethylformamide	100 mL
<b>Biopsy lysis buffer</b>	

10 mM Tris/HCl pH 8.0 (1 M stock)	2 mL
50 mM KCl (1 M stock)	10 mL
0.5 % NP-40 (10 % stock)	10 mL
0.5 % Tween (10 % stock)	10 mL
ad ddH <sub>2</sub> O	200 mL
<b>Colloidal Coomassie G-250 staining solution</b>	
0.02 % Coomassie Brilliant Blue - G250	0.4 g
5 % aluminium-sulfate-(14-18)-hydrate	25 g
10 % Ethanol (96 %)	50 mL
2 % orthophosphoric acid (85 %)	11.75 mL
ad ddH <sub>2</sub> O	500 mL
<b>Flow wash buffer</b>	
D-PBS	450 mL
FCS	50 mL
Sodium azide	0.1 mL
<b>LB medium</b>	
1 % Tryptone	10 g
0.5 % Yeast extract	5 g
0.5 NaCl	5 g
ad ddH <sub>2</sub> O	1 L
<b>LB plates (Ampicillin)</b>	
LB medium	500 mL
1.5 % Agar	7.5 g
100 µg/mL Ampicillin (20 mg/ml)	2.5 mL
<b>Microinjection buffer T<sub>10</sub>E<sub>0.1</sub></b>	
5 mM Tris-base (1 M)	40.85 µL
5 mM TRIS-acid (1 M)	209 µL
0.1 mM EDTA (5mM)	500 µL
Embryo-tested water	24.25 mL
<b>NBT stock solution (10 x)</b>	
1 % Nitrotetrazolium Blue chloride	0.5 g
0.1 M Tris-HCl, pH 9.5	50 mL
<b>TAE buffer (50 x) for agarose gels</b>	
2 M Tris	243.3 g
50 mM EDTA	14.6 g
ad ddH <sub>2</sub> O,, pH 8.5	1 L
	(used as 0.5 x TAE buffer for agarose gels)
<b>Transfer buffer (semi-dry)</b>	
25 mM Tris	3.04 g
19.2 mM Glycine	14.4 g
20 % Isopropanol	200 mL
ad ddH <sub>2</sub> O, pH 8.3	1 L
<b>Tris-Tricin anode buffer A (5 x)</b>	

0.2 M Tris	24.23 g
ad ddH <sub>2</sub> O, pH 8.9	1 L
<b>Tris-Tricin cathode buffer C (5 x)</b>	
0.5 M Tris	60.57 g
0.5 M Tricin	89.59 g
0.5 % SDS	5 g
ad ddH <sub>2</sub> O, pH 8.6	1 L
<b>Tris-Tricin gel buffer (3 x)</b>	
3 M Tris	90.86 g
0.3 % SDS	0.75 g
ad ddH <sub>2</sub> O, pH 8.45	250 mL
<b>Tris-Tricin running gel (10 %)</b>	
10 % Acrylamid	10 mL
1 x Tris-Tricin gel buffer (3 x)	10 mL
ddH <sub>2</sub> O	6.2 mL
1 % Glycerol (87 %)	3.65 mL
0.05 % TEMED	15 µL
0.05 % APS stock solution (10 %)	150 µL
<b>Tris-Tricin sample buffer (4 x)</b>	
0.2 M Tris	1.21 g
48 % Glycerol	24 mL
16 % SDS	8 g
0.4 M DTT	3.085 g
0.04 % Coomassie-Brilliant-Blue R-250	0.02 g
ad ddH <sub>2</sub> O, pH 6.8	50 mL
<b>Tris-Tricin stacking gel</b>	
4 % Acrylamid	2 mL
1 x Tris-Tricin gel buffer (3 x)	5 mL
ddH <sub>2</sub> O	7.9 mL
0.067 % TEMED	10 µL
0.067 % APS stock solution (10 %)	100 µL

**Table M 6:** List of ingredients for self-made buffers.

## 5. Consumable material

Material	Company
96-well plate round bottom, microtest plate	Sarstedt, Nürnberg, Germany
Acclaim PepMap100 C18.5 µm, 100 Å	LC Packings, Thermo Fisher Scientific, Darmstadt, Germany
Butterfly needle	Dahlhausen, Köln; Germany
Cell strainer (0.45 µm, 100 µm)	Corning Incorporated (Falcon), Tewksbury, MA, USA

Centrifugal filter Ultrafree, PTFE	Merck Millipore, Darmstadt, Germany
Cryotubes 1.8 mL	Nunc, Roskilde, Denmark
CytoOne TC treated 48-well plates	StarLab, Hamburg, Germany
EDTA-Monovettes	Sarstedt, Nürnberg, Germany
Erlenmeyer flask, baffled	Schott, neoLab, Heidelberg, Germany
Falcon tubes (15 and 50 mL)	Corning Incorporated (Falcon), Tewksbury, MA, USA
Flow cytometry tubes	Corning Incorporated (Falcon), Tewksbury, MA, USA
HisTrap excel column (Ni-Sepharose)	GE Healthcare, Freiburg, Germany
LoBind tubes	Eppendorf, Hamburg, Germany
NuPAGE Novex Bis-Tris Protein Gels	NuPAGE Novex, Life Technologies, Thermo Fisher Scientific, Darmstadt, Germany
PepMap, 25 cm, 75 µm ID, 2µm/100 Å pore size	LC Packings, Thermo Fisher Scientific, Darmstadt, Germany
Pipette tips	Eppendorf, Hamburg, Germany
Pipettes (1, 2, 5, 10 and 25 mL)	Greiner Bio-One, Frickenhausen; Germany
Protein-A-Sepharose 4 Fast Flow columns	GE Healthcare, Freiburg, Germany
PVDF membrane Immobilon	Merck Millipore, Darmstadt, Germany
QiaShredder Columns	Qiagen, Hilden, Germany
qPCR plates 384 well	Thermo Fisher Scientific, Darmstadt, Germany
Reaction tubes (0.2, 0.5, 1.5 and 2 mL), sterile	Sarstedt, Nürnberg, Germany
Syringe (50 mL)	Braun, Melsungen; Germany
Tissue culture flasks (T25, T75, T175)	Greiner bio-one, Frickenhausen, Germany
Tissue Culture Plates, 24-well flat bottom	Corning Incorporated (Falcon) (#351147), Tewksbury, MA, USA
Tissue Culture Plates, 6-well flat bottom	Corning Incorporated (Falcon), Tewksbury, MA, USA
Whatman filter paper	Nunc, Roskilde, Denmark

**Table M 7:** List of consumable material.

## 6. Instruments

Instrument	Company
Agilent 2100 Bioanalyzer	Agilent Technologies, Santa Clara, CA, USA)
Äkta pure 25L1	GE Healthcare, Freiburg, Germany
autoMACS Pro Separator	Miltenyi Biotech, Bergisch-Gladbach, Germany
Automatic pipette Easypet	Eppendorf, Hamburg, Germany
BD LSR Fortessa	BD Biosciences, Heidelberg, Germany
Centrifuge 541R	Eppendorf, Hamburg, Germany
Centrifuge 5424R	Eppendorf, Hamburg, Germany
Centrifuge Megafuge 1R	Heraeus, Hanau, Germany
ChemoCam Imager for ECL Western Blot detection	Intas, Göttingen, Germany
Eppendorf BioPhotometer	Eppendorf, Hamburg, Germany
Eppendorf Centrifuge 5424 R	Eppendorf, Hamburg, Germany

Eppendorf Centrifuge 5810R	Eppendorf, Hamburg, Germany
Eppendorf MasterCycler pro	Eppendorf, Hamburg, Germany
Eppendorf rotor A-4-81	Eppendorf, Hamburg, Germany
Fraction collector F9-R (Äkta)	GE Healthcare, Freiburg, Germany
Gel imager system	Intas, Göttingen, Germany
Incubator	Heraeus, Hanau, Germany
Innova43 Incubator Shaker	New Brunswick Scientific, Eppendorf, Hamburg, Germany
Liquid nitrogen tank LS6000	Tec-Lab GmbH, Taunusstein, Germany
LKexv 2600 MediLine (+ 4°C)	Liebherr, Biberach an der Riß, Germany
LTQ-Orbitrap XL	Thermo Fisher Scientific, Darmstadt, Germany
Magnetic stirrer RCT basic	IKA Werke, Staufen, Germany
Micro scale	MC1 Research, Sartorius, Göttingen, Germany
Microscope Axiovert 200M	Zeiss, Jena, Germany
Microscope Axiovert 25	Zeiss, Jena, Germany
MilliQ Advantage A10	Merck Millipore, Darmstadt, Germany
Mr. Frosty Freezing Container	Thermo Fisher Scientific, Darmstadt, Germany
Multichannel Pipettes	Eppendorf, Hamburg, Germany
NanoDrop 2000	Thermo Fisher Scientific, Darmstadt, Germany
Neubauer counting chamber	neoLab, Heidelberg, Germany
PCR machine TC-412	Techne Inc, Burlington, NJ, USA
pH meter SenTix 81	WTW GmbH, Weilheim, Germany
Pipettes for disposable tips	Eppendorf, Hamburg, Germany
Power Supply Power Ease 300W	LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Precision balance Kern 770	Kern & Sohn GmbH, Balingen, Germany
Premium Non-Frost, (-20 °C)	Liebherr, Biberach an der Riß, Germany
RSLC, Ultimate 3000	Dionex, Thermo Fisher Scientific, Darmstadt, Germany
Scie-Plas Semi-Dry Blotting Systems	Harvard Apparatus, Holliston, MA, USA
Shaker UnitWist RT	Uni-Equip, Martinsried, Germany
Sterile Hood Thermo Scientific Heraeus Laminar Flow	Thermo Fisher Scientific, Darmstadt, Germany
Thermo Shaker TS100	Kisker Biotech, Steinfurt, Germany
Thermomixer	Eppendorf, Hamburg, Germany
Tube Rotator	VWR International, Darmstadt, Germany
Ultra Low Temperature High Efficiency Freezer U570HEF, (-80 °C)	New Brunswick Scientific, Eppendorf, Hamburg, Germany
ViiA 7 Real Time PCR System	Applied Biosystems, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Vortex Genie 2	Scientific Industries, Bohemia, NY, USA
Waterbath	Julabo, Seelbach, Germany
Xcell I Blot Module	Novex, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany
Xcell Sure Lock Mini-Cell for protein electrophoresis and Western Blotting	Novex, LifeTechnologies, Thermo Fisher Scientific, Darmstadt, Germany

**Table M 8:** List of instruments.

## 7. Software and databases

Name	Company/ source/ reference
Adobe Illustrator CS5	Adobe Systems Incorporated, San Jose, CA, USA
Adobe Photoshop CS5	Adobe Systems Incorporated, San Jose, CA, USA
Antibody Society	<a href="http://www.antibodysociety.org">http://www.antibodysociety.org</a>
BD FACS DIVA 7.0	BD Biosciences, Heidelberg, Germany
BioGrid	<a href="http://thebiogrid.org">http://thebiogrid.org</a> ; <sup>223</sup>
ChemBioDraw Ultra 14.0	PerkinElmer, Waltham, MA, USA
ChemoStar	Intas, Göttingen, Germany
CRISPR Design Tool	<a href="http://crispr.mit.edu/">http://crispr.mit.edu/</a> Zhang Lab, MIT 2015, MA, USA
DrugBank	<a href="http://www.drugbank.ca">http://www.drugbank.ca</a> ; <sup>168</sup>
European Medicines Agency (EMA)	<a href="http://www.ema.europa.eu/ema/">http://www.ema.europa.eu/ema/</a>
Ensembl human database (Release 69; 100607 sequences)	<a href="http://www.ensembl.org/Homo_sapiens/Info/Index">http://www.ensembl.org/Homo_sapiens/Info/Index</a> ; <sup>161</sup>
FlowJo Software	<a href="http://www.flowjo.com">http://www.flowjo.com</a> , Tree Star, Ashland, OR, USA
Generic GO Term Finder	<a href="http://go.princeton.edu/cgi-bin/GOTermFinder">http://go.princeton.edu/cgi-bin/GOTermFinder</a> ; <sup>166</sup>
GeneSpring software GX 12.5	Agilent Technologies, Santa Clara, USA
Google Patents	<a href="http://www.google.com/patents">http://www.google.com/patents</a>
GProx including Mfuzz package	<a href="http://gprox.sourceforge.net">http://gprox.sourceforge.net</a> ; <sup>163,164</sup>
Immunology Link	<a href="http://www.immunologylink.com">http://www.immunologylink.com</a>
LocTree3	<a href="https://rostlab.org/services/loctree2/">https://rostlab.org/services/loctree2/</a> ; <sup>171</sup>
Mascot (Matrix Science, Release number 2.4)	<a href="http://www.matrixscience.com/mascot_support_v2_4.html">http://www.matrixscience.com/mascot_support_v2_4.html</a> Matrix Science Limited, London, UK
Microsoft Office 2011	Microsoft Corporation, Redmond, WA, USA
NCBI Gene Expression Omnibus (GEO)	<a href="http://www.ncbi.nlm.nih.gov/geo/">http://www.ncbi.nlm.nih.gov/geo/</a> ; <sup>169</sup>
PolyPhobius	<a href="http://phobius.sbc.su.se/poly.html">http://phobius.sbc.su.se/poly.html</a> ; <sup>172</sup>
Primer3web	<a href="http://primer3.ut.ee">http://primer3.ut.ee</a> ; <sup>174</sup>
Progenesis LC-MS software (version 4.0)	<a href="http://www.nonlinear.com/progenesis/qi-for-proteomics/">http://www.nonlinear.com/progenesis/qi-for-proteomics/</a> Nonlinear Dynamics, Durham, NC, USA
ProteomeXchange Consortium	<a href="http://www.proteomexchange.org">http://www.proteomexchange.org</a> ; <sup>159</sup>
PubMed	<a href="http://www.ncbi.nlm.nih.gov/pubmed">http://www.ncbi.nlm.nih.gov/pubmed</a>
qPrimerDepot	<a href="https://primerdepot.nci.nih.gov">https://primerdepot.nci.nih.gov</a>
R including gplots package	<a href="https://www.R-project.org/">https://www.R-project.org/</a> ; <sup>173</sup>
REVIGO	<a href="http://revigo.irb.hr">http://revigo.irb.hr</a> ; <sup>167</sup>
UniProtKB/Swiss-Prot	<a href="http://www.uniprot.org">http://www.uniprot.org</a> ; <sup>162</sup>
U.S. Food and Drug Administration (FDA)	<a href="http://www.fda.gov">http://www.fda.gov</a>

**Table M 9:** List of software and databases.

## Table of Figures

- Figure 1: Development and differentiation of the naive CD4<sup>+</sup> T cell and plasticity of T helper cell subsets.
- Figure 2: Balance of the immune system.
- Figure 3: Cell surface glycoprotein labeling via PAL (periodate oxidation and aniline catalyzed oxime ligation) technique.
- Figure 4: Workflow for the generation of the cell surface atlas of human naive and activated CD4<sup>+</sup> T cells.
- Figure 5: Scheme of the PAL-qLC-MS/MS technique.
- Figure 6: Establishment and validation of PAL-qLC-MS/MS.
- Figure 7: Technical monitoring of T cell marker expression during sample preparation for PAL-qLC-MS/MS via flow cytometry.
- Figure 8: Comparability of the donor samples for PAL-qLC-MS/MS.
- Figure 9: Decision Tree for the analysis of qLC-MS/MS-derived data to assemble the cell surface glycoproteome.
- Figure 10: Illustration of the number of identified proteins of the PAL-qLC-MS/MS approach.
- Figure 11: Unsupervised clustering and Gene Ontology enrichment analysis of the cell surface proteins of human naive and activated CD4<sup>+</sup> T cells, identified by PAL-qLC-MS/MS.
- Figure 12: Cell surface proteins exhibiting the highest protein abundance (PAL-q-LC-MS/MS) at the depicted stimulation time points.
- Figure 13: Dynamic expression profiles of selected cell surface proteins not co-cited with T cell biology so far.
- Figure 14: Expression validation of selected cell surface proteins not co-cited with T cell biology so far.
- Figure 15: Unsupervised cluster analysis of the cell surface proteins of human naive and activated CD4<sup>+</sup> T cells, identified by the targeted flow cytometry screening panel.
- Figure 16: Comparison of PAL-qLC-MS/MS and flow cytometry screening panel results.
- Figure 17: Decision Tree for the identification of cell surface protein-coding transcripts from the genome-wide transcriptomic microarray analysis.
- Figure 18: GO term enrichment analysis of transcripts coding for cell surface- and plasma membrane proteins from the genome-wide microarray expression analysis.
- Figure 19: Gene expression analysis of transcripts coding for cell surface- and plasma membrane proteins.
- Figure 20: The surface atlas of human naive and activated CD4<sup>+</sup> T cells.
- Figure 21: Heat map of cell surface proteins showing significantly different regulation in the analysis of the combined omics datasets.
- Figure 22: Correlation of proteomic (PAL-qLC-MS/MS) to transcriptomic results (microarray expression analysis).
- Figure 23: Expression patterns (1) of solute carrier (SLC) transporters obtained via genome-wide microarray analysis (A) and PAL-qLC-MS/MS (B).
- Figure 24: Expression patterns (2) of solute carrier (SLC) transporters obtained via genome-wide microarray analysis (A) and PAL-qLC-MS/MS (B).

- Figure 25: Flow cytometry suitability screen of hybridoma supernatants (containing monoclonal antibodies against c16orf54) for their specificity for murine naive CD4<sup>+</sup> T cells.
- Figure 26: Flow cytometry suitability screen of hybridoma supernatants (containing monoclonal antibodies against c16orf54) for their specificity for human naive CD4<sup>+</sup> T cells.
- Figure 27: Antibody-peptide competition staining.
- Figure 28: Flow cytometry staining of human PBMCs to examine c16orf54 expression on lymphocyte subpopulations.
- Figure 29: Flow cytometry staining of human PBMCs – gated on c16orf54<sup>+</sup> cells.
- Figure 30: Overview of the constructs for recombinant production of c16orf54.
- Figure 31: Western blot analysis of cell lysates of human naive CD4<sup>+</sup> T cells using monoclonal antibodies with specificity for c16orf54 and validation of c16orf54\_4 and c16orf54\_5 expression and purification.
- Figure 32: Strategy for the generation of CRISPR/CAS mediated knockout mice.
- Figure 33: Proof of sgRNA functionality and size-confirmation of T7-PCR templates for in vitro transcription.

## List of Tables

- Table 1: List of therapeutic antibodies currently approved/under regulatory review or in phase II/III clinical trials against cell surface structures on T cells and their secreted products (excluded are antibodies for cancer therapy).
- Table 2: List of all cell surface proteins identified on human naive and/or activated CD4<sup>+</sup> T cells.
- Table 3: Summary of the proteins identified by PAL-qLC-MS/MS, which are not co-cited with T cell biology so far.
- Table 4: Solute carrier (SLCs) transporters identified via PAL-qLC-MS/MS.
- Table 5: Rating of hybridoma supernatant performance in flow cytometry staining (test on murine cells).
- Table 6: Rating of hybridoma supernatant performance in flow cytometry staining (test on human cells).
- Table M 1: List of reagents, media, buffer, enzymes and kits.
- Table M 2: List of antibodies used for T cell stimulation or flow cytometry application.
- Table M 3: List of primary antibodies used for Western blot application.
- Table M 4: List of secondary antibodies used for Western blot application.
- Table M 5: List of primer names and corresponding sequences.
- Table M 6: List of ingredients for self-made buffers.
- Table M 7: List of consumable material.
- Table M 8: List of instruments.
- Table M 9: List of software and databases.

## Abbreviations

Abbreviation	Description
°C	degree Celsius
2D	2-dimensional
A	Ampere
Å	Angstrom
ABC	activated B cell
Abu	alpha butyric acid
AC	accession
AD	atopic dermatitis
ADP	adenosine diphosphate
AHR	aryl hydrocarbon receptor
AIRE	autoimmune regulator
AIT	allergen immunotherapy
AML	acute myeloid leukemia
AP	alkaline phosphatase
APC	antigen presenting cell
APC (antibody conjugate)	allophycocyanin
APS	ammoniumpersulfate
AS	ankylosing spondylitis
ATP	adenosine triphosphate
autoMACS	automated MACS
Bcl-6	B cell lymphoma 6 protein
BCR	B cell receptor
BMMC	bone marrow nuclear cells
bp	base pair
BV	brilliant violet
c16orf54	chromosomal open reading frame 54 on chromosome 16 coding for transmembrane protein c16orf54
Cas9	CRISPR-associated protein (RuvC and HNH type nuclease domains)
CD	cluster of differentiation
cDNA	complementary DNA
CLL	chronic lymphocytic leukemia
CLP	common lymphoid progenitor
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
CSPA	cell surface protein atlas
CTLA-4	cytotoxic T-lymphocyte-associated protein 4
CV	column volume
D	donor
D-PBS	Dulbecco's phosphate buffered saline
Da	Dalton
DC	dendritic cell
DEPC	diethylpyrocarbonate
DLBCL	diffuse large B cell lymphoma
DMEM	Dulbecco's modified eagle medium
DMSO	dimethyl sulfoxide
DN	double negative
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleoside triphosphate
DP	double positive

DSB	Double strand breaks
DTT	dithiothreitol
e.g.	exempli gratia=for example
EAACI	European Academy of Allergy and Clinical Immunology
EB	elution buffer
ECL	enhanced chemiluminescent
EDTA	ethylenediaminetetraacetate
EFNB	eprhin
ELISA	enzyme-linked immunoassay
EMA	European Medicines Agency
ENST	ensemble transcript
ERK	extracellular-signal-regulated kinase
<i>et al.</i>	et alli=and others
EVI	ecotropic viral integration site
Fab	fragment antibody binding
FCS	fetal calf serum
FDA	Food and Drug Administration
FFAR3	free fatty acid receptor 3
Fig.	figure
FITC	fluorescein isothiocyanate
Foxp3	forkhead box P3
FT	flow through
g	gram
g x	g force
G418	Geneticin
GATA	trans-acting T-cell-specific transcription factor
GCB	germinal center B cell
GEO	Gene Expression Omnibus
GFP	green fluorescent protein
GO	Gene ontology
GProx	The Graphical Proteomics Data Explorer
gRNA	guide RNA
h	hour
H+L	heavy and light chain
HEK	human embryonic kidney
HLDA	human leukocyte differentiation antigens
HR	Homologous recombination
HRP	horse radish peroxidase
IBD	inflammatory bowel disease
ICOS	inducible T-cell costimulator
IDG	Institute of Developmental Genetics, Helmholtz Center Munich
IDG	Institute of Developmental Genetics
IFN	interferon
Ig	immunoglobulin
IL	interleukin
ILC	innate lymphoid cell
IMI	Institute of Molecular Immunology
IP	immunoprecipitation
IRF-4	interferon regulatory factor 4
ISH	in situ hybridization
k	kilo
kb	Kilo base pair
KO	knockout
L	liter
lacZ	gene encoding beta-galactosidase

LB	Lysogeny Broth
LC-MS	liquid chromatography – mass spectrometry
LDS (sample buffer)	lithium dodecyl sulfate
m	milli
M	molar
mAB	monoclonal antibody
MACS	magnetic activated cell sorting
MALDI-TOF	matrix-assisted laser desorption/ionization – time of flight
MAP	mitogen-activated protein kinase
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
min	minute
miRNA	micro RNA
MM	multiple myeloma
MOPS	3-Morpholinopropane-1-sulfonic acid
MS	mass spectrometry
ms	multiple sclerosis
n	nano
n/a	not available
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NF1	neurofibromin 1
NHEJ	non-homologous end joining
NK	natural killer
NKT	natural killer T
NMO	neuromyelitis optica
Notch	neurogenic locus notch homolog protein 1
NPTN	neuroplastin
Ova	Ovalbumin
p value	probability
PA	psoriatic arthritis
PAGE	polyacrylamide gel electrophoresis
PAL-qLC-MS/MS	periodate oxidation and aniline-catalyzed oxime ligation coupled to quantitative liquid chromatography-tandem mass spectrometry
PAM	Protospacer adjacent motif
PAMP	pathogen-associated molecular patterns
PBMC	peripheral blood mononuclear cells
PCA	principal component analysis
PCR	polymerase chain reaction
PE	phycoerythrin
Pen/Strep	Penicillin/Streptomycin
PerCPCy5.5	Peridinin chlorophyll protein
pH	potential hydrogen
PI	propidium iodide
PI3-K	phosphatidyl-inositol-3-kinase
PKP	plakophilin
pTreg	peripheral derived regulatory T cell
PVDF	polyvinylidene fluoride
qPCR	quantitative PCR
R	receptor
RA	rheumatoid arthritis
Rag	recombination-activating gene
Ras	rat sarcoma
RAST	radio-allergen-sorbent-test
RefSeq	reference sequence

REVI GO	reduce and visualize gene ontology
RIN	RNA Integrity Number
RNA	ribonucleic acid
RNF	ring finger protein
RORC2	RAR-reöated orphan receptor
RPMI	Roswell Park Memorial Institute medium
RRms	relapsing remitting multiple sclerosis
RSLC	rapid separation liquid chromatography
RT	room temperature
RUNX1/AML1	runt-related transcription factor 1/ acute myeloid leukemia 1 protein
s	second
SAIL	surface antigen in leukemia
SDS	sodium dodecyl sulfate
Sf9	Spodoptera frugiperda
sgRNA	single guideRNA
SLC	solute carrier
SN	supernatant
SP	single positive
STAT	signal transducer and activator of transcription
T-bet	T-box transcription factor TBX21
Tcm	central memory T cell
TCR	T cell receptor
TE	Tris-EDTA buffer
Tem	effector memory T cell
TEMED	tetramethylethylenediamine
Tfh	follicular T helper
TGF- $\beta$	transforming growth factor beta
Th	T helper
T <sub>m</sub>	melting temperature
TM	transmembrane
TMB	3,3',5,5'-Tetramethylbenzidin
TMEM	transmembrane protein
TMH	transmembrane helix
TNFR	tumor necrosis factor receptor
TNF $\alpha$	tumor necrosis factor alpha
Tr1	T regulatory type 1
Treg	regulatory T cell
Tris	Tris(hydroxymethyl)aminomethane
Trm	tissue resident memory T cell
tTreg	thymically derived regulatory T cell
U	unit
UniProt_SL	UniProt Subcellular Localization
UniProtKB	UniProt Knowledge Base
V	Volt
WB	Western Blot
$\mu$	micro

## Acknowledgement

I want to deeply thank Prof. Dr. Benedikt Grothe for being the first reviewer of my dissertation and for representing my dissertation at the Faculty of Biology. Thank you for giving me the great opportunity to accept the HELENA scholarship and to complete the dissertation about this topic externally at the Center of Allergy and Environment.

I want to express my deepest gratitude to Prof. Dr. Carsten Schmidt-Weber for giving me the opportunity to complete my dissertation in his institute. Thank you for your great enthusiasm and support, the possibility to gain experiences at national and international conferences and for the promotion of my entrance into the European Academy for Allergy and Clinical Immunology (EAACI).

Many thanks go to Prof Dr. Angelika Böttger for accepting to be the second reviewer of my dissertation. Many thanks also to Prof Dr. Angelika Böttger and Prof. Dr. Thorsten Buch for being members of my thesis committee and for providing critical suggestions and fruitful discussions during our meetings.

I also want to take this opportunity to thank all other reviewers for taking interest in my work and taking time to review my dissertation.

My deepest thank you goes to Dr. Kathi Suttner, who was my supervisor and mentor for more than 2,5 years during my time as a PhD student. We had a wonderful time together, not only at work. It was a relationship of trust, honesty and teamwork, and also became a wonderful friendship, which I am and will be grateful for the rest of my life. Thank you so much for everything Kathi!

A very deep thank you goes to Dr. Simon Blank who stepped in and became my new supervisor. He did not hesitate to take care of me and supported me with all resources throughout the last months of my time as a PhD student. Thank you so much! An additional important thank you goes to PD Dr. Stefanie Eyerich, who also offered me her full support and helped me a lot in all kinds of situations, especially during the last months as a PhD student.

A special thank you to all collaborating partners of this project, without them this work would not have been possible.

I feel very honored to have received a scholarship for the fully funded PhD student position by the HELENA Graduate School of the Helmholtz Center Munich and for being part of this great Graduate School during my time as a PhD student. Thank you Monika and Gaby for your support from the very beginning.

I am enormously grateful for being able to get to know and to work together with my PhD mates and colleagues Anne, Sarah, Dani, Jenny, Danijel and Dennis. Thank you for your moral and scientific support and your friendship, my time as a PhD student would not have been the same without you. The person who joined my life as PhD student the latest, became one of the most important friends, forever ever thank you Max for helping me in the lab, for reading and discussing my dissertation, for calming me down, for always reminding me about what is important and what isn't, for your shoulder and for simply being there.

I want to specially thank Ferdi and Steffi E. for excellent technical support throughout the whole time. Thank you Ferdi for making it such a great time during endless hours of T cell cloning and thank you Steffi for your help and time without even the need to ask for it.

I want to thank the colleagues of the ZAUM, IAF and the UNIKA-T for the nice atmosphere in our institute: Linda, Natalie, Felix, Kilian, Julia E., Marta, Katharina, Isabelle B., Julia H., Steffi G., Caro, Saskia, Moritz, Francesca W., Jana, Cordula, Elke, Gaby, Susanne, Kristina, Dawn, Jeroen, Christine, Jose, Joana, Gudrun, Renate, Sebastian, Stefan, Caspar, Maria, Mary, Steffi M., Tanja, Johanna, Benjamin, Francesca A.

Many, many thanks to the people I met during my first groundbreaking experiences in a lab, friends who paved my way to research, encouraged me to stay there and built up a wonderful and deep friendship since this time: Katrin aus München, Prakti, Steffi, Krissy, Cookie, Elvir.

Life without friends would not be possible, especially during the time as a PhD student. Thank you for being by my side all of these years, thank you for your enduring support and thank you for helping me to keep the balance. You are simply the best Janine, Anja, Anna and Isabelle.

I am deeply thankful for the help and support of my wonderful sister in law Anne and my parents in law Erika and Friedrich. It is great to know that there are such wonderful people behind you, whenever you need them.

Last, but not least, I want to thank my most important supporters, my husband Michi, my brother Jan and my parents Jutta and Gerhard. Without you and all your support throughout the years, my studies and my dissertation would not have been possible. Thank you for being by my side, for trusting in me and for convincing me that there is always a way to follow. It is an irreplaceable gratitude to have all of you in my life.

## Publications

**Graessel, A.** et al. A Combined Omics Approach to Generate the Surface Atlas of Human Naive CD4+ T Cells during Early T-Cell Receptor Activation. *Mol Cell Proteomics* 14, 2085-2102, doi:10.1074/mcp.M114.045690 (2015).

Aguilar-Pimentel A., **Graessel A.**, Helmut Fuchs, Gailus-Durner V., Hrabe de Angelis M., Blank S., Chaker A., Gutermuth J., Schmidt-Weber C.B. Facilitated tolerance-induction in experimental allergen-specific immunotherapy by inhibition of the JAK1/3 pathway. *Manuscript submitted.*

## Poster & Oral Presentations

### Nov 2013:

Oral presentation at the Allergy Retreat 2013 in Grainau, Germany

Title: "Surface profiling in T cell activation"

### Jan 2014:

Poster presentation at the Meeting of the International Allergy and Immunology Taskforce in Neuherberg, Germany

Title: "Naive T cell surface proteome"

### Feb 2014:

Oral presentation at the 12<sup>th</sup> EAACI Immunology Winter School 2014 in Brasov, Romania

Title: "The surfaceome of naive T cells: Transcriptional and translational library of the naive T cell surface"

Poster presentation at the 7<sup>th</sup> <interact> Symposium 2014 in Munich, Germany

Title: "Transcriptional and translational library of the naive T cell surface - an advanced view on the early activation process"

### Mar 2014:

Oral presentation at the 26<sup>th</sup> DGAKI Annual Allergy Workshop 2014 in Mainz, Germany

Title: "The surface library of naive T cells - an advanced view on the early activation process"

### Sep 2014:

Poster presentation at the 44<sup>th</sup> Annual Meeting of the DGfl 2014 in Bonn, Germany

Title: "Applying a systems biologic approach to identify new cell surface markers on naive T cells potentially involved in the early time window of T cell differentiation"

### Oct 2014:

Poster presentation at the 13<sup>th</sup> Human Proteome Organization World Congress 2014 in Madrid, Spain

Title: "The surfaceome of naive T cells – a proteomic systems approach to identify new cell surface markers in early T cell activation"

### Feb 2015:

Poster presentation at the 13<sup>th</sup> EAACI Immunology Winter School 2015 in Les Arcs 1800, France

Title: "The cell surface atlas of human naïve CD4<sup>+</sup> T cells - reference book for monitoring T cell activation"

**Mar 2015:**

Poster presentation at the WIRM IX 2015 in Davos, Switzerland

Title: "The human naive CD4<sup>+</sup> T cell surface atlas - a reference book for early T cell activation"

**Jun 2015:**

Poster and Oral presentation at the EAACI Congress 2015 in Barcelona, Spain

Title: "Identification of new cell surface proteins on human naive CD4<sup>+</sup> T cells potentially involved in the early T-cell activation"

**Aug 2015:**

Invitation as speaker for the group seminar of AG Dr. Carolin Daniel (Immunological tolerance in Type 1 Diabetes, Institut für Diabetesforschung, Helmholtz Center Munich) in Neuherberg, Germany

Title: "Generation of the cell surface atlas for human naive and activated CD4<sup>+</sup> T cells"

**Feb 2016:**

Poster presentation at the 14<sup>th</sup> EAACI Immunology Winter School 2016 in Cortina d' Ampezzo, Italy

Title: "Solute Carrier (SLC) Transport Proteins Are Diversely Expressed On The Surface Of Human Naive CD4<sup>+</sup> T Cells During Early T Cell Activation"

**Mar 2016:**

Oral presentation at the 28<sup>th</sup> DGAKI Annual Allergy Workshop 2016 in Mainz, Germany

Title: "Solute Carrier (SLC) transport proteins – early triggers during the activation of human naive CD4<sup>+</sup> T cells?"

## Scholarships, Grants and Awards

**Scholarship:** Fully funded PhD student position of HELENA Graduate School (international application round of HELENA 2012 - Helmholtz Center Munich) for 3 years (Jan 2013-Jan 2016)

**Grant:** Travel Grant for the 12<sup>th</sup> EAACI Immunology Winter School 2014 in Brasov, Romania

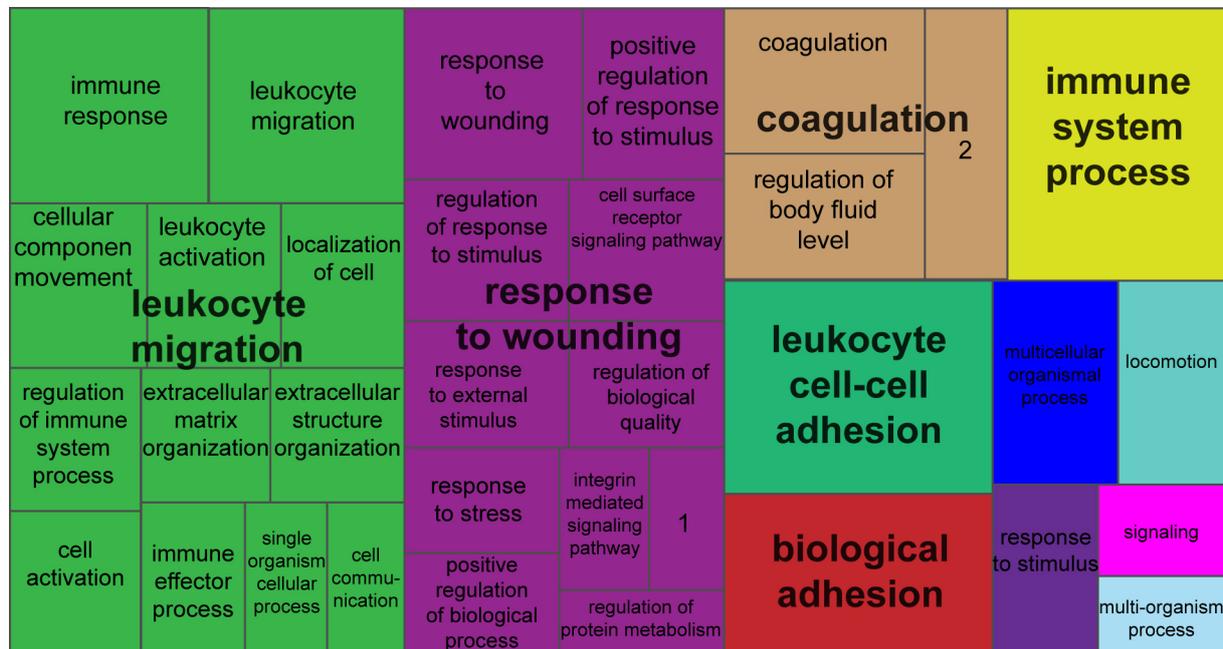
**Grant:** Travel Grant for the 13<sup>th</sup> Human Proteome Organization World Congress 2014 in Madrid, Spain

**Award:** Poster Prize at the 13<sup>th</sup> EAACI Immunology Winter School 2015 in Les Arcs 1800, France

**Scholarship:** Congress Scholarship for the EAACI Congress 2015 in Barcelona, Spain

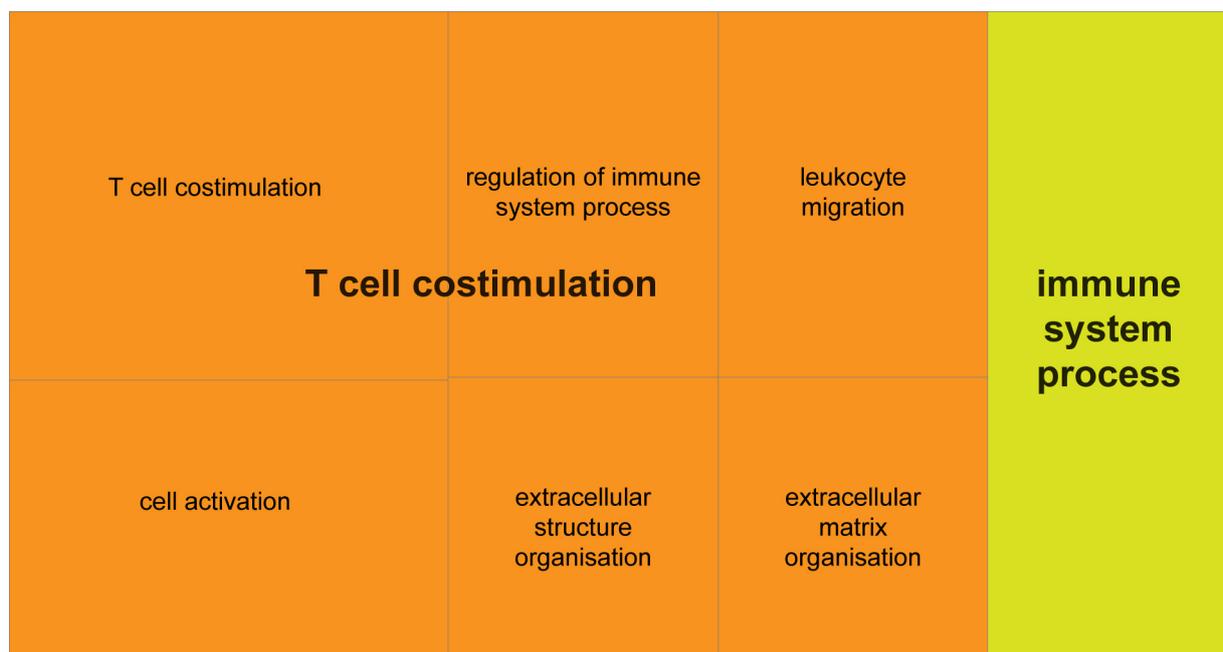
**Award:** Poster Prize at the 14<sup>th</sup> EAACI Immunology Winter School 2016 in Cortina d' Ampezzo, Italy

## Appendix

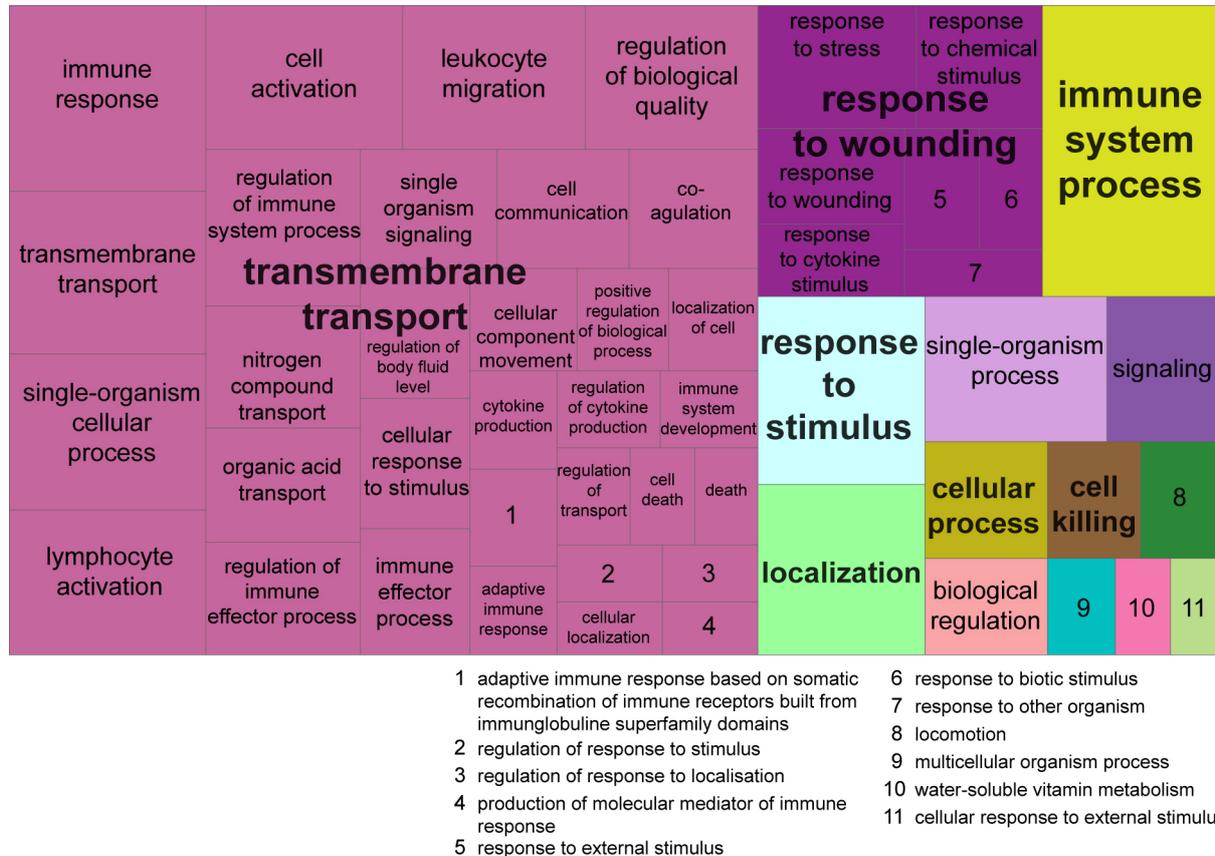


1 regulation of peptidyl-tyrosine phosphorylation 2 single-multicellular organism process

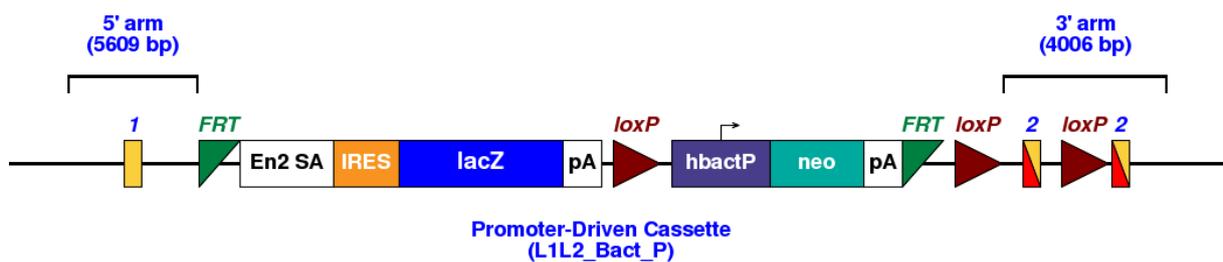
**Figure A 1: REVIGO Treemaps of GO term enrichment analysis of PAL-qLC-MS/MS results – Cluster 1.** Proteins identified via PAL-qLC-MS/MS were grouped into three different GProx clusters and analyzed via GO term enrichment analysis (GProx membership value  $\geq 0.6$ ). The GO terms (biological processes) were summarized by using REVIGO and visualized as treemaps. Each rectangle can be seen as a single representative of an enriched GO term. The single representatives are joined into “superclusters” of terms, which are loosely related, shown in different colors. The adjusted size of the rectangles reflects the p-value of the GO term in the underlying GOA database.<sup>157</sup>



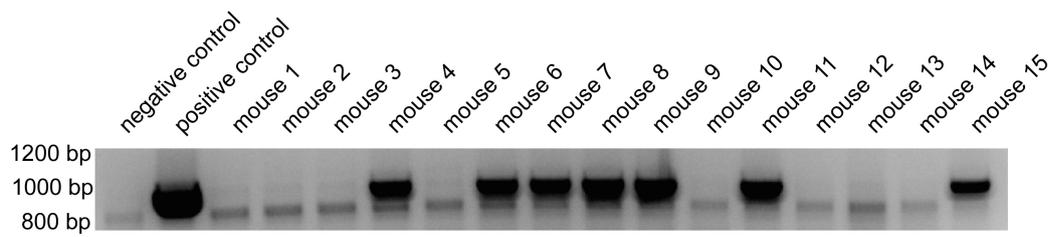
**Figure A 2: REVIGO Treemaps of GO term enrichment analysis of PAL-qLC-MS/MS results – Cluster 2.** See figure text of Figure A1.<sup>157</sup>



**Figure A 3: REVIGO Treemaps of GO term enrichment analysis of PAL-qLC-MS/MS results – Cluster 3.** See figure text of Figure A1.<sup>157</sup>



**Figure A 4: Targeting vector for the production of AI467606\_KOfirst.** Sperm from mice, which were genetically modified with this targeting vector for the generation of the mutant mouse line AI467606\_KOfirst, was purchased from Baylor College, Houston, TX, USA. (1=5' UTR, FRT=flippase (FLP) recognition site, En2A=splice acceptor, IRES=internal ribosomal entry site, lacZ=lacZ reporter gene, pA=polyadenylation signal, loxP=Cre recombinase acceptor site, hBactP=autonomous promoter, neo=neomycin resistance gene, 2=coding exon) (© 2016 IMPC · International Mouse Phenotyping Consortium)



**Figure A 5: Agarose gel of lacZ genotyping of AI467606\_KOfirst mice.** The products of the genotyping PCR for the 15 different mice of the line AI467606\_KOfirst (including one negative and one positive control) were loaded on an agarose gel to check which of the mice carries the recombined sequence of the targeting vector. The band with the correct size for the lacZ amplicon is 950 bp long. The Mice 4, 6, 7, 8, 9, 11, and 15 showed the mutant genotype for the mouse line AI467606\_KOfirst.

## Supplement

The supplemental figures and tables are recorded on the attached DVD.

**Supplemental Figure S1:** Spectra of manual annotation of one-peptide hits.<sup>157</sup>

**Supplemental Table S1:** PAL-qLC-MS/MS: Peptide identification for Trypsin and PNGase F digested samples.<sup>157</sup>

**Supplemental Table S2:** PAL-qLC-MS/MS: Protein identification for Trypsin and PNGase F digested samples.<sup>157</sup>

**Supplemental Table S3:** Bioinformatics analysis of the transcriptomic dataset.<sup>157</sup>

**Supplemental Table S4:** GProX membership values of all proteins identified via PAL-qLC-MS/MS to the respective clusters.<sup>157</sup>

**Supplemental Table S5:** GProX membership values of all proteins identified via flow cytometry surface screening to the respective clusters.<sup>157</sup>