

A Combined Omics Approach to Generate the Surface Atlas of Human Naive CD4⁺ T Cells during Early T-Cell Receptor Activation^{*}

Anke Graessel[‡], Stefanie M. Hauck[§], Christine von Toerne[§], Edda Kloppmann[¶], Tatyana Goldberg^{¶**}, Herwig Koppensteiner^{‡‡}, Michael Schindler^{‡§§}, Bettina Knapp^{¶¶}, Linda Krause^{¶¶}, Katharina Dietz^{‡|||}, Carsten B. Schmidt-Weber^{‡|||}, and Kathrin Suttner^{‡|||}

Naive CD4⁺ T cells are the common precursors of multiple effector and memory T-cell subsets and possess a high plasticity in terms of differentiation potential. This stem-cell-like character is important for cell therapies aiming at regeneration of specific immunity. Cell surface proteins are crucial for recognition and response to signals mediated by other cells or environmental changes. Knowledge of cell surface proteins of human naive CD4⁺ T cells and their changes during the early phase of T-cell activation is urgently needed for a guided differentiation of naive T cells and may support the selection of pluripotent cells for cell therapy. Periodate oxidation and aniline-catalyzed oxime ligation technology was applied with subsequent quantitative liquid chromatography-tandem MS to generate a data set describing the surface proteome of primary human naive CD4⁺ T cells and to monitor dynamic changes during the early phase of activation. This led to the identification of 173 N-glycosylated surface proteins. To independently confirm the proteomic data set and to analyze the cell surface by an alternative technique a systematic phenotypic expression analysis of surface antigens via flow cytometry

was performed. This screening expanded the previous data set, resulting in 229 surface proteins, which were expressed on naive unstimulated and activated CD4⁺ T cells. Furthermore, we generated a surface expression atlas based on transcriptome data, experimental annotation, and predicted subcellular localization, and correlated the proteomics result with this transcriptional data set. This extensive surface atlas provides an overall naive CD4⁺ T cell surface resource and will enable future studies aiming at a deeper understanding of mechanisms of T-cell biology allowing the identification of novel immune targets usable for the development of therapeutic treatments. *Molecular & Cellular Proteomics* 14: 10.1074/mcp.M114.045690, 2085–2102, 2015.

Naive CD4⁺ T cells are the common precursors for all other T-helper cell subsets and it is of fundamental importance for specific immunity that their differentiation process is well directed. A complex signaling network is engaged upon antigen recognition that triggers the differentiation process of stem-cell-like, plastic, antigen-unexperienced naive T cells into antigen-specific, functional distinct T-cell subphenotypes (1). The differentiation process of naive T cells is tightly regulated in healthy individuals. Pathology develops under dysregulated effector responses such as overshooting responses leading to impaired tolerance (2) or ineffective control of infections (3). Naive T cells are defined by CD45RA expression and they are early cellular targets of immune modulation regarding the differentiation process and the development of long lasting, sustainable therapeutic strategies. In contrast, memory T cells express CD45RO and cover already committed cells such as T helper 1 and T helper 2 cells. Therefore, we chose to investigate the naive CD4⁺ T cell (CD45RA) and its phenotype during T-cell receptor (TCR)¹ activation. The differentiation

From the [‡]Center of Allergy and Environment (ZAUM), Technische Universität und Helmholtz Zentrum München, Munich, Germany; [§]Research Unit Protein Science, Helmholtz Zentrum München, Neuherberg, Germany; [¶]Department of Informatics, Bioinformatics & Computational Biology i12, Technische Universität München, Garching/Munich, Germany; ^{||}New York Consortium on Membrane Protein Structure (NYCOMPS), New York Structural Biology Center, New York, New York 10027; ^{**}TUM Graduate School, Center of Doctoral Studies in Informatics and its Applications (CeDoSIA), Technische Universität München, Munich, Germany; ^{‡‡}Institute of Virology, Helmholtz Zentrum München, Neuherberg, Germany; ^{§§}Institute of Medical Virology and Epidemiology of Viral Diseases, University Clinic Tübingen, Tübingen, Germany; ^{¶¶}Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; ^{|||}DZL- Member of the German Lung Research Center

Received October 23, 2014, and in revised form, May 13, 2015

Published, MCP Papers in Press, May 19, 2015, DOI 10.1074/mcp.M114.045690

Author's contribution: AG, SMH, MS, CBSW, KS: designed experiments; AG, CvT, HK, KD, KS: performed experiments; AG, SMH, CvT, EK, TG, BK, LK, KS: analyzed data; CBSW, KS: supervised the project; SMH, CvT: contributed to manuscript preparation; AG, KS: wrote the final version of the manuscript.

¹ The abbreviations used are: TCR, T-cell receptor; α CD3/ α CD28, antiCD3 and antiCD28; CD, cluster of differentiation; GO, Gene Ontology; PAL, periodate oxidation and aniline-catalyzed oxime ligation; PBMC, peripheral blood mononuclear cell; RIN, RNA integrity number; SLC family, solute carrier family; TMH, transmembrane helix; UniProt_SL, UniProt subcellular localization.

process of naive CD4⁺ T cells is initiated by ligand binding to the TCR, costimulatory surface receptors, and co-acting of specific extracellular signals and growth factors. This complex interaction, including signals mediated by other cells or changes in the environment, allows the integration of complex immunological conditions.

Until now, approaches dealing with T-cell differentiation focused mainly on genome-wide transcriptome and epigenome investigations revealing a large number of potential key drivers important in T-cell commitment (4–6). However, proteomic approaches dealing with the T-cell differentiation are rarely performed but consistently requested by the immunological community (7, 8). In 2014, two mass-spectrometry-based drafts of the complete human proteome were published on the same day in the same journal highlighting the importance and the need of proteomic data (9, 10). The first proteomic manuscript regarding activated human primary T helper cells, published in 2001, consisted of 91 proteins identified by metabolic labeling, 2-dimensional gel electrophoresis, and MALDI-TOF MS (11). Most of the already existing studies regarding T-cell biology are often conducted in Jurkat T-cell lines instead of primary T cells, focusing on proteomic events during activation close to the TCR, located in lipid rafts (12–14). Other studies focused on T-cell subproteomes within the early stages of T-cell differentiation and investigated proteomic changes in the nucleus of activated human cord blood CD4⁺ T cells after interleukin-4 stimulation (15) or focused on changes of the global phosphoproteome of human primary T cells in response to 5 min of TCR activation with α CD3 (16). *In vitro* manipulated T cells were previously analyzed such as 7-day cultures of *in vitro* differentiated T helper 1 and T helper 2 cells (17), however, the surface proteome of human naive CD4⁺ T cells and how these proteins change during the early time window of α CD3/ α CD28 activation has not been investigated so far. Interestingly, knowledge of the surface proteins of naive CD4⁺ T cells would provide a molecular fingerprint to classify naive CD4⁺ T cells and especially their cellular state during activation. Furthermore, the investigation of cell surface proteins and their expression changes during stimulation might lead to the identification of interesting targets for therapeutic approaches such as antibody-based drugs or enrichment strategies.

The main aim of the present study was to generate an extensive *ex vivo* surface atlas of human naive CD4⁺ T cells and to analyze surface proteins of freshly isolated cells or directly upon stimulation on a deep proteomic level. Therefore an omics approach was performed, combining quantitative proteomic and transcriptomic methods. For the identification of surface proteins of CD4⁺/CD45RA⁺ naive T cells by mass-spectrometry the recently described PAL (periodate oxidation and aniline-catalyzed oxime ligation) approach (18) was applied and adapted for the use with primary T cells and combined with quantitative LC-MS/MS (19). This method allows an efficient labeling of cell surface sialic acid-containing gly-

cans on living cells. Similar technical cell surfaceome approaches were conducted in human and mouse studies before and already showed the importance of these cell surface focusing technologies in order to identify proteins via shotgun proteomics. Most studies are based on the cell surface capture technology, which covalently labels extracellular glycan moieties on living cells (20). These studies focused for example on the surface proteome of stem cells (21), mesenchymal stromal cells (22), and murine adipocytes in obesity (23) as well as other immune cells such as B cell lines derived from lymphomas (24).

Furthermore, a systematic expression analysis of 332 surface antigens via flow cytometry was performed and the combination of the two proteomic data sets (PAL-qLC-MS/MS and flow cytometry) resulted in a cell surface protein atlas of human naive CD4⁺ T cells containing 229 cell surface proteins.

Additionally, a cell surface expression atlas based on transcriptomic data was created to increase the resolution of the proteomic data set. A bioinformatics analysis was performed on the transcriptome of naive and activated CD4⁺ T cells and annotations with experimental evidence or prediction of subcellular localization were used to select genes coding for cell surface proteins, *i.e.* transmembrane proteins localized in the plasma membrane and other proteins localized on the cell surface. This uncovers an additional set of genes encoding proteins located on the surface of human naive CD4⁺ T cells that have to be further validated and might include interesting candidates potentially applicable for a guided T-cell differentiation.

The here described cell surface atlas provides novel potential immune targets usable for therapeutic treatments and also gives for the first time an extensive overview of the cell surface expression pattern of CD markers and surface molecules detected on human naive and activated CD4⁺ T cells.

EXPERIMENTAL PROCEDURES

Cell Isolation and Stimulation—Peripheral blood mononuclear cells (PBMCs) were isolated from the heparinized blood of healthy, nontopic blood donors ($n = 19$) by standard lymphoprep (Fresenius Kabi Norge AS for Axis-Shield PoC AS, Oslo, Norway) density gradient centrifugation. To generate and validate the surface atlas of human naive CD4⁺ T cells on a proteomic level, 15 different biological replicates were used. For the generation of the transcriptome data set by RNA microarray, additional four different donors were included. The study was approved by the local ethics committee of the Technical University Munich, ethic board no 2877/10. Informed written consent was obtained from all study subjects. Naive CD4⁺ T cells were separated by negative isolation (Naive CD4⁺ T cell Kit II, Miltenyi Biotec, Bergisch Gladbach, Germany), followed by the depletion of CD45RO⁺ cells (Isolation Kit CD45RO Microbeads; Miltenyi Biotec) according to the manufacturer's protocol reaching a purity of at least 98%, as monitored by flow cytometry. Cells were activated with plate-bound α CD3 (0.75 μ g/24-well culture plate well; BD Biosciences, Heidelberg, Germany) and soluble α CD28 (0.75 μ g/ml; BD Biosciences) or the T-cell activation/expansion kit with a bead-to-cell ratio of 1:2 (Miltenyi Biotec) in a density of 1×10^6 cells/ml of AIMV

Medium (Gibco by Life Technologies, Darmstadt, Germany) and cultured at 37 °C in a humidified 5% CO₂ atmosphere.

Cell Surface Labeling, Protein Digestion, and N-glycopeptide Enrichment—Periodate oxidation and aniline-catalyzed oxime ligation (PAL)-based cell surface labeling procedure was performed as described previously (18) and adapted for primary human naive CD4⁺ T cells. All following steps were performed on ice. The cell pellet of 8 × 10⁶ naive CD4⁺ T cells per donor (*n* = 4, donors 1–4) and time point was washed twice in ice-cold labeling buffer (PBS with CaCl₂/MgCl₂, pH 6.7) and then added to 1 ml oxidation/biotinylation mix in an one-pot reaction, consisting of 1 mM NaIO₄, 100 μM aminooxy-biotin (Biotium Inc. Hayward, CA) and 10 mM aniline (Sigma-Aldrich, Taufkirchen, Germany) in labeling buffer and incubated for 30 min at 4 °C in the dark, rotating. The biotinylation reaction was quenched by adding glycerol to a final concentration of 1 mM for additional 5 min on the rotator at 4 °C. Cells were washed with 1 ml cold washing buffer (PBS with CaCl₂/MgCl₂, pH 7.4) and an aliquot of cells was collected to check biotinylation labeling efficiency (supplemental Fig. S1A). After centrifugation (4000 rpm, 4 °C, 10 min), each cell pellet was resuspended in 250 μl lysis buffer (1% Nonidet P-40, 10 mM NaCl, 10 mM Tris, pH 7.6, 2 × EDTA-free complete protease inhibitor mixture (Roche, Basel, Switzerland) in ddH₂O) and frozen at –20 °C. To prepare cell membrane fractions, the lysed cells were thawed on ice and constantly vortexed. Raw lysates were cleared by centrifugation (6000 × *g*, 4 °C, 10 min). The pellet was discarded and the supernatant was diluted 1:5 with washing buffer. The diluted samples were incubated with 60 μl of prewashed Strep-Tactin Superflow 50% suspension (IBA GmbH, Göttingen, Germany) in LoBind tubes (Eppendorf, Hamburg, Germany) at 4 °C for 2 h on a rotator to bind biotinylated cell surface proteins to high-affinity streptavidin beads. Samples were centrifuged at 1000 × *g* for 1 min to pellet the bead-protein complexes. The following washing and incubation steps (25) were all carried out in a volume of 200 μl and the centrifugation steps during the washing procedure were carried out at 2000 × *g* for 2 min. First, beads were washed with the washing buffer, supplemented with 0.2% Nonidet P-40 and followed by a washing step with 0.5% SDS (Gibco by Life Technologies) in washing buffer. After centrifugation, the beads of each sample were incubated with 0.5% SDS and 100 mM DTT in washing buffer at room temperature for 30 min. Beads were then washed with UC buffer (6 M urea, 100 mM Tris-HCl, pH 8.5). For alkylation the beads were incubated in UC buffer containing 50 mM iodoacetamide (Sigma-Aldrich) at room temperature for 30 min, followed by washing steps with UC buffer, 5 M NaCl, 100 mM Na₂CO₃ (Merck, Darmstadt, Germany), pH 11.5, and 50 mM Tris-HCl, pH 8.5. Bead-protein complexes were then first digested in 40 μl 50 mM Tris-HCl, pH 8.5, containing 1 μg sequencing-grade modified trypsin (Promega, Madison, WI) in a Thermomixer (Eppendorf) at 37 °C overnight. The samples were centrifuged as before and the supernatant, containing the tryptic peptides, was transferred into a new LoBind tube (Eppendorf). Beads were resuspended in 40 μl 50 mM Tris-HCl, pH 8.5, and centrifuged, the resulting supernatant was then pooled with the first tryptic fraction. The beads were washed with 40 μl 1 × G7 buffer (NEB, Frankfurt a.M., Germany) and then incubated with 20 μl 1 × G7 buffer containing 500 U glycerol-free PNGase F (NEB) at 37 °C in a thermomixer for 6 h to release the glycopeptides from the beads. After the second digest, beads were centrifuged and the supernatant was transferred into a new LoBind tube. The beads were again resuspended with 20 μl 1 × G7 buffer and centrifuged to pool the resulting supernatant with the first PNGase F fraction. Tryptic and PNGase F peptide fractions were stored at –20 °C until mass spectrometric analysis and measured separately.

As a technical validation of the PAL-qLC-MS/MS approach, a flow cytometry analysis was performed with identical samples in parallel to the MS setting, using specific antibodies for several selected surface

markers (CD11a-FITC, CD69-PE, CD62L-PE-TexasRed) (BD Biosciences). The comparison of the expression patterns obtained by flow cytometry or mass spectrometry showed a very high concordance between the two techniques (supplemental Fig. S1B).

LC-MS/MS—The LC-MS/MS analysis were performed as described previously (26, 27) on a LTQ-Orbitrap XL (Thermo Fisher Scientific, Waltham, MA) with the following adjustments: A nano trap column was used (300 μm inner diameter × 5 mm, packed with Acclaim PepMap100 C18, 5 μm, 100 Å; LC Packings, Sunnyvale, CA) before separation by reversed phase chromatography (PepMap, 25 cm, 75 μm ID, 2 μm/100 Å pore size, LC Packings) operated on a RSLC (Ultimate 3000, Dionex, Sunnyvale, CA) using 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 5 min and an equilibration for 15 min to starting conditions. From the MS prescan, the 10 most abundant peptide ions were selected for fragmentation in the linear ion trap if they exceeded an intensity of at least 200 counts and were at least doubly charged. During fragment analysis, high-resolution (60,000 full-width half maximum) MS spectra were acquired in the Orbitrap with a mass range from 300 to 1500 Da. One microscan was recorded with fill times in the FT (MS) set to 0.5 s and in the Ion Trap (MSn) to 0.1 s. Automatic Gain Control (AGC) targets were set to 1e⁺6 (MS) and 1e⁺4 MSn, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (28) via the PRIDE partner repository with the data set identifier PXD001432.

Database Search, Label-Free Relative Quantification, Data Processing, and Identification of Cell Surface Proteins—The RAW files (Thermo Fisher Scientific) were further analyzed using the Progenesis LC-MS software (version 4.0, Nonlinear Dynamics, Newcastle, UK), as described previously (19), with the following changes: For peptide identification, using the search engine Mascot (Matrix Science, Release number 2.4), 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, methionine oxidation, and asparagine or glutamine deamidation were allowed as variable modifications. Spectra were searched against the Ensembl human database (Release 69; 100,607 sequences) (29) excluding the common contaminants keratin and albumin. A Mascot-integrated decoy database search using the Percolator algorithm calculated an average peptide false discovery rate of < 1% when searches were performed with a Percolator score cutoff of 13 and a significance threshold of *p* < 0.05. Search results and spectral files have been uploaded to the ProteomeXchange platform (<http://www.proteomexchange.org>) and are available with the identifier PXD001432. Peptide assignments were re-imported into Progenesis LC-MS. Normalized abundances of all unique peptides were summed up and allocated to the respective protein (supplemental Tables S1 and S2).

To generate the cell surface protein data set with high confidence, proteins were only pursued further if one of the following criteria was true: (1) proteins were identified either in Trypsin or PNGase fraction with one or more peptides if the confidence score was ≥ 18, and (2) protein was identified in both, the Trypsin and PNGase fraction with at least one peptide and a confidence score ≥ 13. For this analysis, the Ensembl human database protein ID of each identified protein was converted to the respective transcript (ENST) ID to guarantee stringency. All spectra from single peptide hits were manually inspected and a complete collection of spectra for the one-peptide hits was added as supplemental Fig. S2.

Furthermore, to verify the identification of cell surface proteins and to remove further contaminants, we considered only proteins for the cell surface protein atlas, which were at least annotated as “membrane” or “secreted” in the UniProtKB/Swiss-Prot database (30). The

experimental evidence for their localization on the cell surface is brought by the PAL-qLC-MS/MS approach. In addition, proteins already nomenclatured as CD molecules were also included in further analysis [supplemental Fig. S3](#) displays a Venn diagram and gives an overview of the number of proteins identified in the trypsin fraction, in the PNGase F fraction or in both fractions. Furthermore, the subcellular localizations of the identified proteins, which are given by the UniProtKB/Swiss-Prot database, are listed.

Flow Cytometry—Surface staining of human naive CD4⁺ T cells ($n = 3$, donors 5–7) was performed with the LEGENDScreen Human Cell Screening (PE) Kit (Biolegend, San Diego, CA) according to the manufacturer's protocol. The kit contains 332 PE-conjugated monoclonal antibodies against human cell surface markers, plus 10 mouse, rat, and hamster Ig isotype controls. Acquisition was performed with the BD LSR Fortessa and BD FACSDIVA 7.0 and data analysis was performed using BD FACSDIVA 7.0 (BD Bioscience) and FlowJo Software (Tree Star, Ashland, OR). To distinguish positive from negative antibody signals for the cell surface proteins on naive and activated CD4⁺ T cells, the mean fluorescence intensity (MFI) detection threshold for all donors and time points was set to the highest measured Ig isotype control. Additionally, positive signals had to be obtained in at least two donors (otherwise stated) to integrate the respective proteins into the cell surface protein atlas.

Furthermore, flow cytometry analysis was also used to check cell purity and viability after isolation and to investigate labeling efficiency and potential side effects of the one-pot reaction on protein expression ([supplemental Fig. S1A and S1B](#)). To this end, 100,000 cells/staining were washed once with flow washing buffer (PBS containing 5% Hyclone FCS (PERBIO, Lausanne, Switzerland) and 0.02% NaN₃) and stained with the respective antibodies (CD4-APC-Cy7, CD3-PerCPCy5.5, CD25-PE-Cy7, CD45RA-Horizon V450, CD45RO-Alexa700, CD69-PE, CD62L-PE-TexasRed, Streptavidin-PE) (BD Biosciences) for 30 min at 4 °C in the dark, washed again with flow washing buffer and analyzed with BD LSRFortessa (BD Biosciences) in combination with the software BD FACSDIVA 7.0 (BD Biosciences). Propidium iodide positive cells were excluded from the analysis.

Unsupervised Clustering by GProX and Gene Ontology Enrichment Analysis—The mean ratios of the via PAL-qLC-MS/MS identified protein abundances as well as the mean fluorescence intensity ratio of the proteins identified via the flow cytometry screening panel (of one representative donor) were subjected to unsupervised clustering (GProX) based on the fuzzy c-means algorithm as implemented in the Mfuzz package (31, 32).

Based on the PAL-qLC-MS/MS clustering results, a Gene Ontology (GO) (33) enrichment analysis was performed: proteins with a GProX membership value ≥ 0.6 to the respective cluster were subjected to the Generic Gene Ontology (GO) TermFinder software (34) and the results were transferred to REVIGO for reduction and visualization. Within the REVIGO software, we set the allowed similarity to the predefined value “small” in order to guarantee that possible pairs of GO terms will have a semantic similarity less than 0.5 (35).

Whole-genome Microarray Analysis and In Silico Selection of Cell Surface Proteins—RNA from human naive CD4⁺ T cells and α CD3/ α CD28 stimulated (3 h) T cells of four blood donors ($n = 4$, donors 8–11) was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) and afterward RNA quality was proofed using an Agilent 2100 Bioanalyzer with the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA) obtaining RNA integrity numbers (RIN). 25 μ g of total RNA (RIN ≥ 9) was amplified and Cy3-labeled using the 1-color Low Input Quick Amp Labeling Kit (Agilent Technologies), according to the manufacturer's protocol. Hybridization to SurePrint G3 Human Gene Expression 8 $\times$ 60K microarray was performed by using the Hybridization Kit (Agilent Technologies). The array data were deposited in NCBI's Gene Ex-

pression Omnibus (36) and are accessible through GEO Series accession number GSE61983.

Quality control was performed with the GeneSpring software GX 12.5 (Agilent Technologies). Only genes, which were detected in four of four blood donors either in the medium control, in the stimulated state, or in both conditions were chosen for further analysis.

Using this selection criteria of the genome-wide scan, genes coding for cell surface proteins localized in and close to the plasma membrane were identified by a bioinformatics analysis. Of a total of 27,958 Entrez Gene RNAs, 17,757 microarray probe names were assigned to 14,455 unique NCBI RefSeq (37) accession numbers by the manufacturer. We could map 13,028 of these RefSeq accession numbers to their corresponding human UniProtKB (30) accession numbers (AC; UniProt release 2013_10). In general, several human UniProt ACs can be assigned to the same gene name. To reduce redundancy, for each gene name we chose, if available, the reviewed UniProt ACs (Swiss-Prot) and otherwise the unreviewed (TrEMBL). Thereby, we assigned UniProt ACs to 12,263 gene names ([supplemental Table S3](#)). (Remaining redundancy is because of 29 gene names that are assigned to more than one reviewed UniProt AC. Additionally, we retained 36 unreviewed UniProt ACs that we predicted to be plasma membrane proteins (criteria see below)). For these UniProtKB ACs, we extracted the subcellular localization (UniProt_SL) annotation from the UniProtKB/Swiss-Prot database (30), if available. Otherwise, we predicted subcellular localization using LocTree3 (38) and transmembrane helices (TMHs) using PolyPhobius (39). We identified cell surface proteins as follows ([supplemental Table S3](#)): (1) From UniProt_SL, we accepted all experimentally verified and probable subcellular localizations. Here, we were in particular interested in cell surface and cell membrane (plasma membrane) annotations. For proteins localized in the plasma membrane, we additionally required that they are single- or multi-pass membrane proteins. If no further information is given, we additionally required at least one TMH predicted by PolyPhobius, otherwise these were classified as putative cell surface proteins. For peripheral or lipid-anchored membrane proteins, we required that they be localized on the extracellular side of the plasma membrane. Those located on the cytoplasmic side were excluded; otherwise, these proteins were classified as putative cell surface proteins. (2) If we could not annotate the protein using UniProt_SL, we used the subcellular localization predicted by LocTree3 and required at least one TMH predicted by PolyPhobius for plasma membrane proteins.

Genes passing these selection criteria were analyzed for differential gene expression with the GeneSpring software GX 12.5 (Agilent Technologies) using the paired Student *t* test, filtered for a corrected *p* value ($p \leq 0.05$, Benjamini-Hochberg correction).

For combining the omics data sets, we first scaled the flow cytometry and the PAL-qLC-MS/MS measurements by subtracting the mean from each value and dividing by the standard deviation. As a result, both protein data sets were comparable with the normalized log₂ expression values of the microarray data. On the scaled data a one-sample, two-sided Welch *t* test was applied to determine significantly differentially expressed proteins/genes at each time point and/or for each technology with a *p* value ≤ 0.01 and a scaled absolute expression measurement ≥ 1 . For this combined data analysis, the R programming language (www.r-project.org) and the “gplots” package were used.

Targeted Validation of Cell Surface Proteins by Real-time PCR and Western Blot Analysis—Isolated total RNA (RNeasy Mini-Kit, Qiagen) from human naive CD4⁺ T cells ($n = 4$, donors 12–15) was subjected to reverse transcription using a high-capacity cDNA kit (Applied Biosystems, LifeTechnologies), following the manufacturer's instructions. Real-time PCR was performed by using the FastStart Universal SYBR Green Mastermix (Roche) and the ViiA 7 Real-Time PCR Sys-

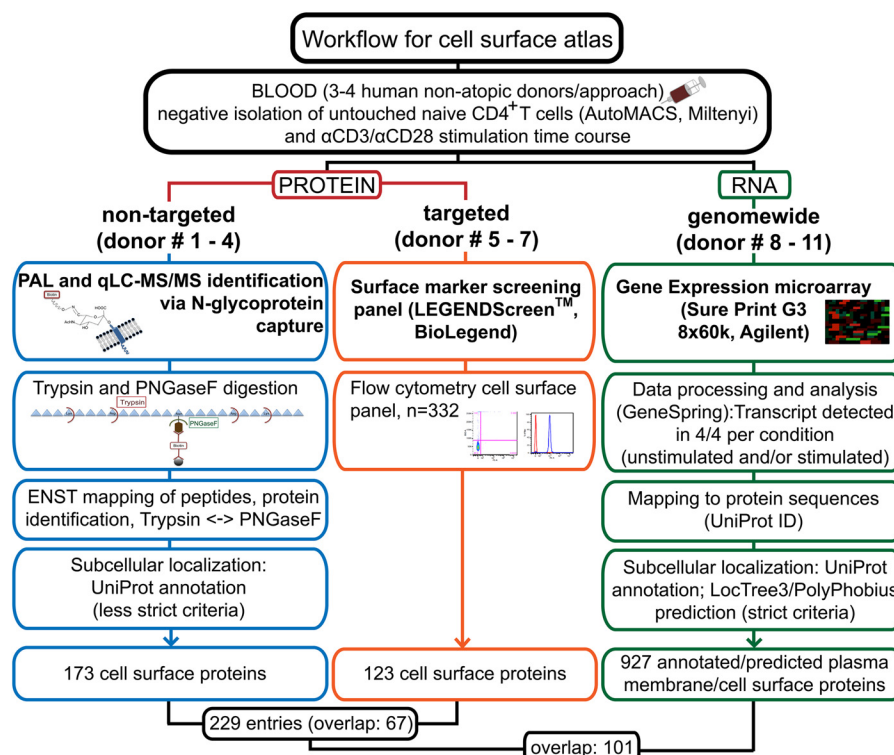


FIG. 1. Workflow to generate the cell surface atlas of human naive and activated CD4⁺ T cells. This scheme combines the generation of samples and summarizes two proteomic-based techniques and one transcriptomic technique, which were conducted to create an extensive cell surface atlas of human naive and activated CD4⁺ T cells. Naive CD4⁺ T cells were isolated from 3–4 different nonatopic human blood donors per technique and activated via αCD3/αCD28 treatment in a time course experiment. The different steps of sample/data processing are listed and the number of identified proteins is given at the end of each technical track. The overlap of the individual outcomes between the different techniques is also shown.

tem (Applied Biosystems, Life Technologies). The specific primers (Metabion, Munich, Germany) used in the real-time PCR are listed in [supplemental Table S4](#). All amplifications were carried out at least in technical duplicates.

For Western blot analysis, equal protein amounts of isolated naive CD4⁺ T cells from four additional blood donors ($n = 4$, donors 16–19) in lysis buffer (1% Nonidet P-40, 10 mM NaCl, 10 mM Tris, pH 7.6, 2 × EDTA-free complete protease inhibitor mixture (Roche) in ddH₂O) were mixed with NuPAGE LDS Sample Buffer (Novex, Life Technologies, Thermo Fisher Scientific, Waltham, MA) and boiled at 95 °C for 10 min, loaded on 10% Bis-Tris protein gels (NuPAGE Novex, Life Technologies) and separated for 1.5 h at 120 V using SDS-PAGE gelelectrophoresis (XCell SureLock Mini-Cell) in a MOPS SDS running buffer (NuPAGE Novex, Life Technologies). Afterward the proteins were transferred to a polyvinylidene difluoride (PVDF) membrane (Merck Millipore, Darmstadt, Germany) using XCell I Blot Module (Novex, Life Technologies) at 60 V (limited to 500 mA) for 90 min. Incubation in 3–5% nonfat dry milk in 1 × PBS for one hour was done for blocking of unspecific binding sites. The membranes were then separately incubated with the primary antibodies ([supplemental Table S5](#)) in 3–5% nonfat dry milk in 1 × PBS overnight at 4 °C. Membranes were washed three times with 3–5% nonfat dry milk in 1 × PBS and incubated with the corresponding HRP-linked secondary antibody ([supplemental Table S5](#)) in 3–5% nonfat dry milk in 1 × PBS for 2 h at 4 °C. Membranes were washed three times with 1 × PBS containing 0.02% Tween and then incubated with the substrate for HRP (Amersham Biosciences ECL Prime Western blotting Detection Reagent, GE Healthcare, Pittsburgh, PA) for 5 min. Chemiluminescence on all Western blots was recorded by ECL ChemoCam Imager and

ChemoStar software (INTAS, Göttingen, Germany). A Western blot assay was declared as positive if there was detection of a band at the expected molecular weight of a protein.

RESULTS

Cell Surface N-Glycoproteome of Human Naive CD4⁺ T Cells—For the identification of the N-glyco surface proteome, human naive CD4⁺ T cells from four biological replicates were isolated and the previously described PAL (periodate oxidation and aniline-catalyzed oxime ligation)-based cell surface protein labeling and enrichment technology (18) was used with subsequent quantitative LC-MS/MS (PAL-qLC-MS/MS) (Fig. 1, blue track). Applying the stringent identification threshold described in the methods section, 242 proteins expressed on naive and activated CD4⁺ T cells were identified and based on the subcellular localization (UniProt_SL) annotation given by the UniProtKB/Swiss-Prot database, 173 of them were included in the cell surface protein atlas (a complete list with these identified proteins is presented in Table I). Remarkably, for 131 of the 173 identified proteins the UniProt_SL annotation for “plasma membrane,” “(cell) membrane,” or “secreted” was given with experimental evidence or an affiliation as CD molecule was already assigned ([supplemental Fig. S3](#)). For the remaining 42 of the proteins, our data set

TABLE I
Cell Surface Protein Atlas of human naive and activated CD4⁺ 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	ASMUH3	2	▲●	a
CD3G (2)		ENST00000532917	P09693	2	▲●	a
CD4	T4	ENST00000116533	P01730	6	▲●	a
CD5	LEU1	ENST00000347785	P06127	17	▲●	a
CD6	T12	ENST00000313421	P30203	9	▲●	a
CD7	gp40	ENST00000312648	P09564	4	▲●	a
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	▲●	a
CD28	TP44	ENST00000324106	P10747	3	▲●	a
CD29	ITGB1	ENST00000302278	P05556	3	▲●	a
CD31	PECAM1	ENST00000563924	P16284	6	▲●	a
CD37	TSPAN26	ENST00000323906	P11049	1	▲●	a
CD38	ADP-ribosyl cyclase	ENST00000228279	P28907	4	▲●	a
CD43	SPN	ENST00000360121	P16150	4	▲●	a
CD44	LHR	ENST00000263398	P16070-12	13	▲●	a
CD45	PTPRC	ENST00000352140	E9PC28	58	▲●	a
CD45RA	PTPRC				▲●	a
CD45RB	PTPRC				▲●	a
CD45RO					▲●	a
CD46 (1)	MCP	ENST00000322875	P15529-2	5	▲●	1)
CD46 (2)		ENST00000358170	P15529	3	▲●	1)
CD47	MER6	ENST00000355354	Q08722-3	3	▲●	a
CD48	BCM1	ENST00000368046	P09326	21	▲●	a
CD49d	ITGA4	ENST00000233573	P13612	1	▲●	a
CD49e	ITGA5	ENST00000293379	P08648	1	▲●	a
CD49f	ITGA6	P23229			▲●	a
CD50	ICAM3	ENST00000160262	P32942	6	▲●	a
CD52	CAMPATH-1	P31358			▲●	a
CD53	MOX44	P19397			▲●	a
CD54	ICAM1	ENST00000264832	P05362	10	▲●	a
CD55	DAF	ENST00000314754	P08174	2	▲●	a
CD58	LFA-3	ENST00000369487	P19256	1	▲●	a
CD59	Protectin	ENST00000351554	P13987	3	▲●	a
CD61	ITGB3	P05106			▲●	a
CD62L	SELL	ENST00000236147	P14151-2	4	▲●	a
CD63	MLA1	P08962			▲●	a
CD69	CLEC2C	ENST00000228434	Q07108	5	▲●	a
CD71	TFRC	ENST00000360110	P02786	18	▲●	a
CD73	NTSE	P21589			▲●	a
CD74	IGLALG	ENST00000009530	P04233	1	▲●	i
CD79b	IGB	P40259			▲●	a
CD81	TAPA1	P60033			▲●	a
CD82	KAI1	ENST00000227155	P27701	6	▲●	a
CD83		Q01151			▲●	a
CD84	SLAMF5	ENST00000311224	Q9UIB8	3	▲●	a
CD85j	LILRB1	Q8NHL6			▲●	a
CD85k	LILRB4	Q8NHJ6			▲●	a
CD95	FAS (1)	ENST00000352159	H7C528	1	▲●	a
	FAS (2)	ENST00000355740	P25445	3	▲●	a
CD96	T-cell surface protein tactile	P40200			▲●	a
CD97		ENST00000242786	P48960	11	▲●	a
CD98	SLC3A2	ENST00000338663	P08195-2	26	▲●	a
CD98LC	SLC7A5	ENST00000261622	Q01650	2	▲●	a
CD99	MIC2	ENST00000381187	P14209-3	1	▲●	a
CD100	SEMA4D	ENST00000356444	Q92854	8	▲●	a
CD101	IGSF2	Q93033			▲●	a
CD102	ICAM2	ENST00000418105	P13598	6	▲●	a
CD105	ENG	ENST00000344849	P17813-2	2	▲●	a
CD107a	LAMP1	ENST00000332556	P11279	4	▲●	a
CD108	SEMA7A	ENST00000261918	O75326	2	▲●	a
CD109	CPAMD7	ENST00000287097	Q6YHK3	2	▲●	a
CD116	CSF2RA	P15509			▲●	a
CD120b	TNFRSF1B	ENST00000376259	P20333	2	▲●	a
CD126	IL6R	P08887			▲●	a
CD127	IL7Rα	P16871			▲●	a
CD129	IL9R	Q01113			▲●	a
CD131	IL2RB	P32927			▲●	a
CD132	IL2RG	ENST00000374202	P31786	1	▲●	a
CD134	TNFRSF4	ENST00000379236	P43489	2	▲●	a
CD137	TNFRSF9	Q07011			▲●	a
CD138	SDC1	P18827			▲●	a
CD144	CDH5	P33151			▲●	a
CD147	BSG	ENST00000333511	P35613	15	▲●	a
CD148	PTPRJ	ENST00000418331	Q12913	1	▲●	a
CD150	SLAMF1	ENST00000235739	Q13291	4	▲●	a
CD152	CTLA4	P16410			▲●	a
CD153	TNFSF8	ENST00000223795	P32971	2	▲●	a
CD154	CD40LG	ENST00000370628	Q3L8U2	1	▲●	a
CD156c	ADAM10	ENST00000260408	O14672	2	▲●	i
CD158aH	KIR2DL1	Q8IS14			▲●	a
CD158f	KIR2DL5A	Q8N109			▲●	a
CD162	SELPLG	ENST00000228463	Q14242-2	2	▲●	a
CD164	MUC-24	Q04900			▲●	a
CD165	SLC38A5	ENST00000317669	Q8WUX1	2	▲●	a
CD166	ALCAM	ENST00000306107	Q13740	2	▲●	a
CD172g	SIRPG	Q9P1W8			▲●	a
CD183	CXCR3	P49682			▲●	a
CD184	CXCR4	P61073			▲●	a
CD196	CCR6	P51684			▲●	a
CD197	CCR7	P32248			▲●	a
CD200	MOX1	ENST00000315711	P41217-2	3	▲●	a
CD200R	CD200R1	Q8TD46			▲●	a
CD205	Ly75	O60449			▲●	a
CD218a	IL18R1	ENST00000233957	Q13478	2	▲●	a
CD225	DNAM-1	ENST00000280200	Q15782	2	▲●	a
CD229	LY9	ENST00000263285	Q9H8G7	1	▲●	a
CD230	PRNP	ENST00000379440	P04156	1	▲●	a
CD235ab	GYPA	P02724			▲●	a
CD236	GYPC	ENST00000259254	P04921	1	▲●	a
CD243	ABCB1	ENST00000265724	P08183	1	▲●	a
CD245 (p220/240)	Dy12, Dy35	Q14207, Q99973			▲●	a
CD257	TNFSF13B	Q9Y275			▲●	a
CD258	TNFSF14	ENST00000245912	Q43557	1	▲●	a

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

TABLE I—continued

[illegible]

expands the hitherto existing knowledge and provides direct experimental evidence for their plasma membrane-associated localization.

To further characterize the obtained protein data set, an extensive NCBI PubMed literature and a further UniProtKB/Swiss-Prot information search were performed. Confirming the robustness of our approach, most of the identified 173 proteins were described in the context of activation/proliferation of T cells or other immune cells (86%). Within this literature search, for 24 of the identified proteins no cocitation with “T cell and/or activation/proliferation/differentiation” could be found (Table I). To validate the expression of these proteins, naive CD4⁺ T cells were collected from four additional donors (donors 12–15) and by performing real-time PCR, all membrane anchored proteins of this group were confirmed at the transcriptional level ([supplemental Fig. S4A](#)). Validated antibodies were available for four identified targets (EVI2A, NPTN, RNF149, TMEM2). Using samples from four additional different biological replicates (donors 16–19) the expression of these markers at protein level could be additionally confirmed ([supplemental Fig. S4B](#)) by an alternative technique (Western blot analysis). The protein expression of the remaining markers has to be proven once the respective antibodies are commercially obtainable.

Unsupervised Clustering and Gene Ontology (GO) Enrichment Analysis of N-glycoproteins During the Early Time Window of TCR Activation—To investigate the changes in protein expression levels on the surface of human naive CD4⁺ T cells occurring during the first hours of activation, quantitative label-free proteomics were applied. To this end, human naive CD4⁺ T cells (donors 1–4) were either left unstimulated or activated by α CD3/ α CD28 for 3, 6, 12, 24, 48 h and subjected to PAL-qLC-MS/MS sample preparation. A principle component analysis showed highly concordant protein abundances between the different blood donors at the respective time-points ([supplemental Fig. S5](#)). The large majority of proteins changed their proteomic expression within the first three hours of activation and/or at a later time point of 24 h of stimulation. Six proteins (CD98LC, CD120b, CD218a, CD258, CD272, CD357) could not be detected on the surface of unstimulated naive CD4⁺ T cells. Ten proteins (SBSN, DAG1, Fas-(2), HLA-B-(3), RNF149, S1PR4, SLC1A4, SLC6A6, TNFRSF18, TNFSF8) showed a differential expression on the naive CD4⁺ T cells among the different donors, based on the PAL-qLC-MS/MS data set, but were expressed during the stimulation at later time-points.

Unsupervised clustering (Fig. 2A) divided the 173 proteins, which were identified on the surface of human naive and activated CD4⁺ T cells according to their dynamic changes in protein abundance into three clusters ([supplemental Table S6](#)): Cluster 1 consists of 40 proteins and Cluster 2 contains 25 proteins. Both clusters show a fast expression decrease within the first hours of activation. Whereas the expression pattern increases in Cluster 1 at later time points, the expres-

sion in Cluster 2 stays at a lower level over time and does not reach or exceed the initial expression value. Cluster 3 is the largest one and contains 108 proteins. The expression of these proteins does only slightly change during the first hours of activation, however shows a strong up-regulation after 24 h of activation.

To obtain deeper insight into which kind of proteins are present in the different expression clusters and what they do have in common, a Gene Ontology (GO) Enrichment Analysis was performed. The Generic GO TermFinder algorithm was applied to identify enriched GO terms, shared among the proteins in each cluster. To reduce the obtained list with enriched GO terms, REVIGO was used to summarize and visualize the GO term results as treemaps (Fig. 2B, for more information see [supplemental Fig. S6](#)). As expected, some GO terms are similar between all clusters such as immune system process or response to stimulus, however, there are also very specific hits for each cluster. Proteins contained in Cluster 1 are primarily involved in migration, adhesion, and response to wounding. Cluster 2, characterized by a quick expression change within the first hours of activation, is enriched for proteins responsible for T-cell costimulation and activation. Cluster 3 covers proteins showing an increase in their expression at a later time point and is especially enriched for transmembrane transport. Fig. 2C shows the dynamic expression profiles of selected proteins that are involved in various processes enriched in the preceding GO term analysis. Certainly, the depicted proteins are associated in general to more than one GO term, but were selected as representatives for one GO term supercluster. The graphs at the bottom do not refer to any GO term supercluster, but illustrate the expression changes of those proteins during activation, which were not cocited in the context of T-cell biology so far.

Extension and Validation of the Surface Proteome of Naive CD4⁺ T Cells Via a Flow Cytometry Surface Screen—In order to independently confirm the data obtained by mass spectrometry and to establish the cell surface atlas of naive CD4⁺ T cells by an alternative technique, human naive CD4⁺ T cells were subjected to an immunostaining with 332 different PE-labeled monoclonal antibodies against known cell surface proteins (Fig. 1, orange track). Because the MS-based data clearly showed that most of the proteins have their strongest expression changes either at an early (3 h) or a later (24 h) time point, the flow cytometry analysis was restricted to these time-points. Using a commercially available cell surface screening kit, 123 markers could be detected on the surface of naive and/or activated CD4⁺ T cells (Table I). According to their expression changes within three and 24 h of α CD3/ α CD28 stimulation, these surface markers could be divided into three clusters ([supplemental Table S7](#)), in agreement with the results obtained via the PAL-qLC-MS/MS approach (Fig. 3A). Fig. 3B displays histograms of the flow cytometric analysis of numerous selected cell surface markers of each cluster

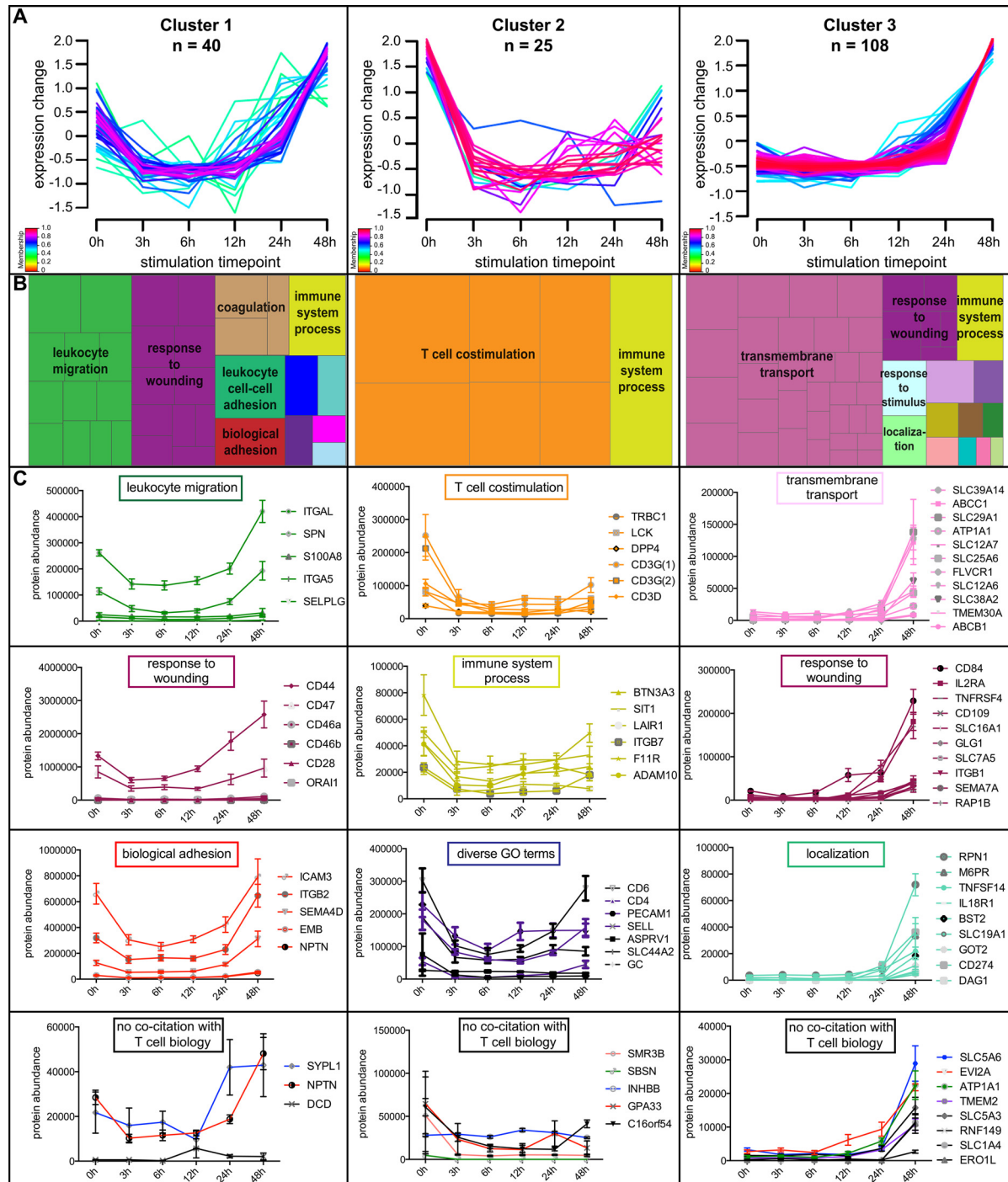


FIG. 2. Unsupervised clustering and Gene Ontology enrichment analysis of the cell surface N-glycoproteome of human naive CD4⁺ T cells identified via PAL-qLC-MS/MS. A, GProX cluster analysis of all via PAL-qLC-MS/MS identified proteins ($n = 173$) according to their expression profile during α CD3/ α CD28 stimulation (3, 6, 12, 24, and 48 h). The membership value represents how well a protein fits into the respective cluster. Proteins with a membership value smaller than 0.4 are not shown (a summary of all membership values is given in [supplemental Table S6](#)). B, Gene Ontology (GO) term analysis visualized by REVIGO treemaps for proteins contained in the respective clusters with a GProX membership value of at least 0.6. Each rectangle is a single representative of an enriched GO term. The representatives are joined into "superclusters" of loosely related terms, visualized with different colors. Size of the rectangles are adjusted to reflect the p value of the GO term in the underlying GOA database (35). See [supplemental Fig. S6](#) for detailed description of the smaller rectangles. C, Dynamic expression profiles (including standard error of the mean of four biological replicates) of selected proteins during α CD3/ α CD28 stimulation (3, 6, 12, 24, and 48 h). The proteins are grouped according to the preceding GO term analysis. The lowest panel, labeled "no cocitation with T-cell biology," highlights the dynamic expression profiles of selected proteins, which are not cocited in the context of T-cell biology so far.

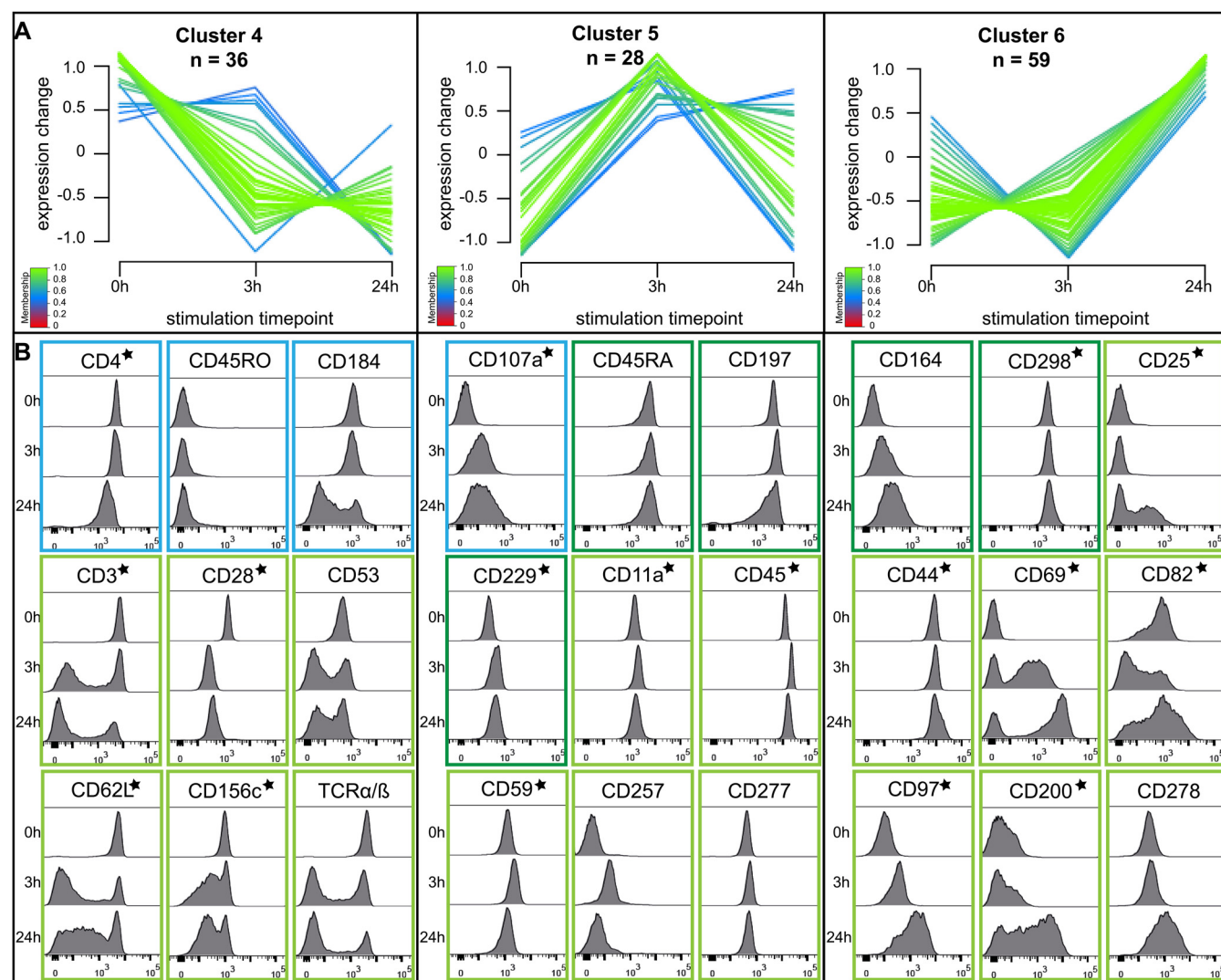


FIG. 3. Unsupervised clustering and expression profiles of the cell surface markers of human naive CD4⁺ T cells identified via the targeted flow cytometry screening panel. A, GProX cluster analysis of all via flow cytometry identified proteins ($n = 123$) according to their expression profile (MFI) during α CD3/ α CD28 stimulation (3 and 24 h) of one representative donor. The membership value represents how well a protein fits into the respective cluster. Proteins with a membership smaller than 0.4 are not shown (a summary of all membership values is given in supplemental Table S7). B, Histograms of the flow cytometric analysis of several selected cell surface markers, which were detected on naive and/or activated CD4⁺ T cells, representing each GProX cluster. The protein expression changes of one representative donor during α CD3/ α CD28 stimulation (3 and 24 h) are displayed. The color frame of each histogram illustrates the membership value of the protein to the respective GProX cluster (blue: membership 0.4–0.6; dark green: 0.6–0.8, light green: 0.8–0.99). Surface markers that were also identified in the PAL-qLC-MS/MS approach are labeled with an asterisk.

and illustrates the expression profiles during activation in detail.

The comparison between the PAL-qLC-MS/MS and the flow cytometry analysis showed that a set of 56 proteins was uniquely detected by flow cytometry-based cell surface phenotyping. A disagreement between PAL-qLC-MS/MS technology and flow cytometry was only observed for five proteins, which were identified by PAL-qLC-MS/MS, but not detectable by the antibody in the flow cytometric profiling at any examined time point. Markers which were not detected in the flow cytometry-based surface screen

might be potential negative markers for naive CD4⁺ T cells ($n = 209$).

Collectively, the combination of the mass spectrometry-based data set and the targeted flow cytometry screening panel expanded the knowledge of proteins present on the cell surface. This permits the generation of an extensive surface protein atlas of human naive and activated CD4⁺ T cells containing in total 229 surface proteins (a summary of all identified proteins can be found in Table I).

In silico Identification of Cell Surface Proteins and Integration of the Transcriptomic Data Set into the Proteomic-based

Cell Surface Atlas—The combination of the nontargeted PAL-qLC-MS/MS approach and the targeted antibody-based cell surface screen led to the generation of a cell surface protein atlas. Because of technical reasons, PAL-qLC-MS/MS offers in general a hypothesis-free and system-wide analysis but is limited to the identification of N-glycosylated surface proteins and low abundant proteins are often difficult to detect. On the other hand, the flow cytometry-based surface screen cannot be used to identify novel and uncharacterized surface markers because a predefined, commercially available antibody panel against known surface markers was used.

To circumvent these technical limitations and to identify additional proteins, which were not covered by the two proteomic approaches, a surface expression atlas relying on transcriptomic data, subcellular annotations based on experimental evidence or predicted subcellular localization, was created (Fig. 1, green track). For that purpose, human naive CD4⁺ T cells of four biological replicates (donors 8–11) were either left unstimulated or exposed to α CD3/ α CD28 for 3 h to investigate changes in gene expression during the early time window of TCR activation and subjected to a whole-genome microarray analysis. Genes, coding for cell surface proteins were identified by a bioinformatics analysis among detected genes from the genome-wide scan. In particular, to specifically identify cell surface proteins, we used (1) experimentally verified or (2) probable UniProtKB subcellular localization annotations and, if not available, (3) predicted the subcellular localization (LocTree3) and TMHs (PolyPhobius) ([supplemental Table S3](#)). This strict selection process led to the identification of 927 genes coding for cell surface proteins. 141 of these genes display a significant expression change within the first three hours of activation ($FC \geq 1.5$, $p \leq 0.05$ corrected) (Fig. 4).

A comparison between the surface proteins identified in the proteomic approaches (PAL-qLC-MS/MS and flow cytometry-based surface screen) and the genes coding for surface proteins, identified via bioinformatics analysis, showed that 53% of the proteins identified by mass-spectrometry and flow cytometry analysis could also be found in the transcriptional surface expression data set. To further deepen our analysis, we generated in addition a putative cell surface data set, in which we extenuated the bioinformatics selection criteria. Here, we included proteins that are described in the UniProtKB database as potentially expressed on the extracellular side of the plasma membrane, e.g. lipid-anchored or peripheral membrane proteins for which it is not known whether they are on the extracellular side (Table I and [supplemental Table S3](#)). This putative data set presents 248 additional genes, potentially encoding cell surface proteins. Using this putative cell surface data set for the comparison with the proteomic approach, we reached an overlap of 58% between transcriptomic and proteomic data sets. The overlap could even be improved to 83%, if proteins, which are solely mentioned as “membrane” proteins in the UniProtKB (no more information

regarding their subcellular localization), are considered in the analysis.

Fig. 5 depicts the combined proteomic and transcriptomic data sets and compares the quantitative expression trends of all detected cell surface proteins, which were identified via PAL-qLC-MS/MS as well as flow cytometry or detected either by PAL-qLC-MS/MS or flow cytometry alone, with the respective mRNA expression data (given mRNA data is independent of the annotation or prediction as a cell surface protein of the respective gene). A large set of cell surface proteins showed a weak correlation between expression patterns at the transcriptional compared with the protein level at the early time points. Of note, Fig. 5 indicates a reference, presenting a scheme of numerous surface and CD markers detected on naive CD4⁺ T cells and their corresponding changes during T-cell activation, combining transcriptome and proteome technologies.

In order to select out of this reference map the most promising surface markers for further functional studies, we combined the transcriptomic and the proteomic data set on the level of significant regulation. We identified a distinct group of 32 cell surface markers, which were measurable at RNA as well as at protein level (PAL-qLC-MS/MS and/or flow cytometry) and revealed a significant expression change in at least one technical approach for a minimum of one stimulation time point (Fig. 6). Obviously, a large majority of these proteins belong to the cluster of differentiation and thus their role in T-cell biology is already well characterized. However, Fig. 6 highlights one cell surface marker, EVI2A, whose expression was also confirmed by Western blot analysis. For this protein a definite general functional role and, in particular, a role during activation is not described so far and further studies are essential to identify its potential function in the context of T-cell activation.

DISCUSSION

In this study, we present an extensive surface atlas describing proteins located on the surface of human naive and activated CD4⁺ T cells. These cells are the common precursor for all other T helper cell subsets and are therefore of fundamental importance for future developments in immunology and clinical application.

We conducted a combined omics approach to link proteomic and transcriptomic data. First, we combined two complementary proteomic techniques, mass-spectrometry and flow cytometry and could create an extensive cell surface protein atlas of human naive and activated CD4⁺ T cells, containing 229 proteins. All proteins are directly measured on protein level and thus represent an experimentally validated set of cell surface proteins involved in the T-cell activation process. Next, a cell surface expression atlas based on transcriptomic data, annotations based on experimental evidence as well as predicted subcellular localization, was built and identified 927 genes coding for proteins located on the

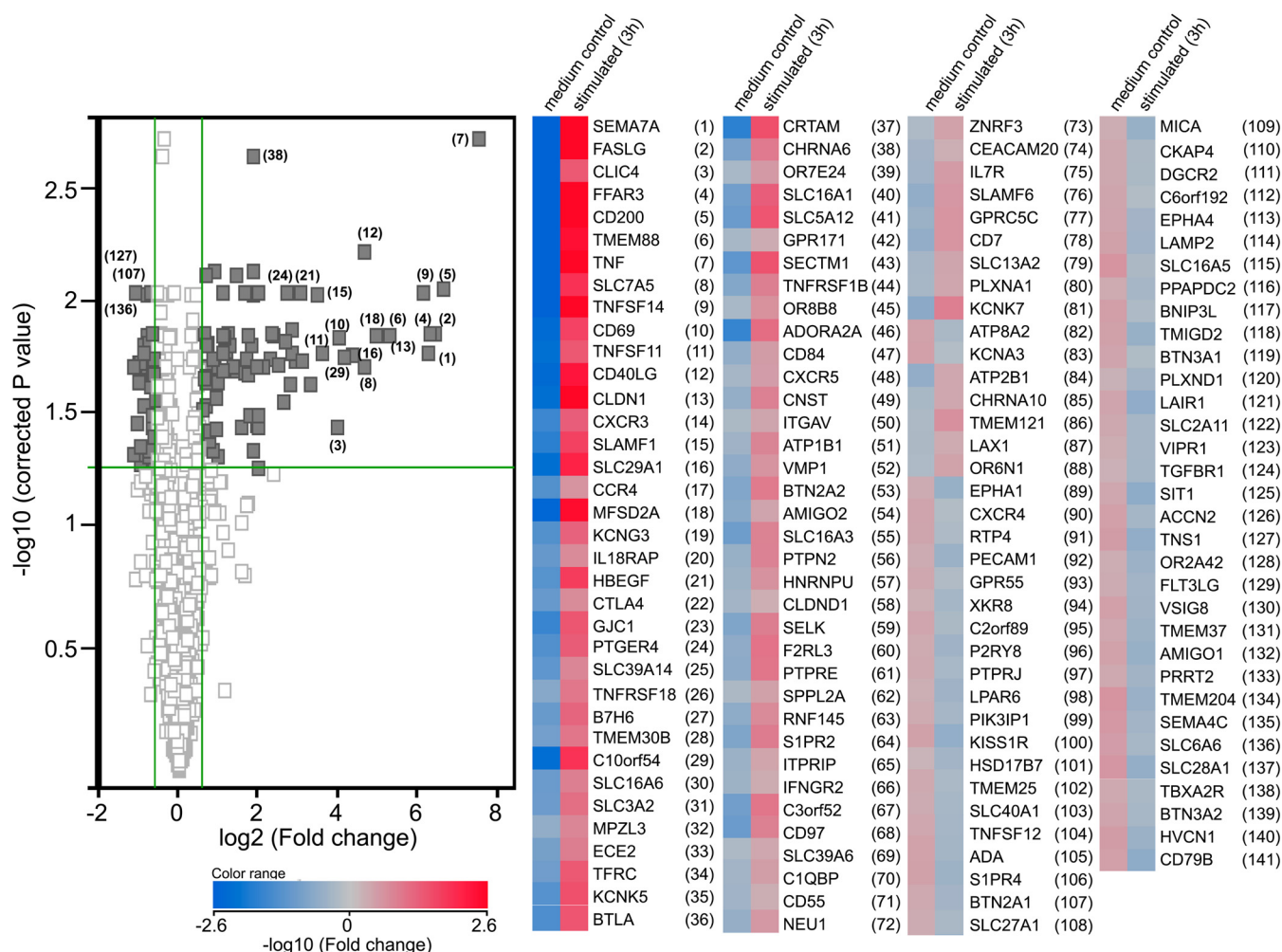


FIG. 4. Gene expression analysis of the transcriptional surface proteome. Based on a whole-genome microarray analysis, cell surface proteins were annotated by experimental UniProtKB subcellular localization, LocTree3, and PolyPhobius predictions. The differential expression of these genes ($n = 927$) in human naive CD4⁺ T cells is shown as volcano plot (left) illustrating the fold change ($\log_2$) compared with p value ($-\log_{10}$, corrected p value) after 3 h of α CD3/ α CD28 stimulation. The horizontal line denotes a corrected significance level of 0.05 and the vertical one denotes an absolute fold change of 1.5.* Among 927 genes annotated or predicted as cell surface proteins, 141 genes showed significant expression changes above the fold change cutoff. The corresponding heat map of these 141 significantly regulated genes is given on the right and represents the fold change values observed between unstimulated (medium control) and stimulated (α CD3/ α CD28 for 3h) naive CD4⁺ T cells. Expression levels are represented in a color scale from blue (low expressed) to red (highly expressed). (*if multiple probes per gene were measured, only the highest value was included in the analysis.)

surface of human naive and activated CD4⁺ T cells, of which 101 could be confirmed in the combined proteomic data set (Table I).

In detail, by using the PAL-qLC-MS/MS approach, we identified 173 N-glycoproteins on the surface of human naive CD4⁺ T cells with high confidence. An extensive literature search revealed that 24 of these proteins were not mentioned in the context of T-cell biology before. We validated all membrane-anchored markers ($n = 20$) via qPCR, however a confirmation on protein level via an alternative technique was only doable for four targets. The proteomic validation of the remaining markers is still missing because of the absence of validated detection tools for these markers, e.g. such as monoclonal antibodies for flow cytometry. Certainly, to spec-

ify these 24 markers as novel T-cell surface proteins would be too broad, because their expression needs to be further substantiated on protein level with an alternative technique. However, very recently, Bausch-Fluck *et al.* published a mass-spectrometry-derived cell surface protein atlas providing a surfaceome snapshot of 78 different human and murine cellular species. Among them a mixed CD4⁺/CD25⁻ T-cell population is listed, which contains nonactivated, naive as well as memory T cells (40). This population can be clearly distinguished from the pure naive CD4⁺/CD45RA⁺ T-cell population, which are investigated in the current manuscript, and has no further focus on the activation process. However, 84 proteins identified by our PAL-qLC-MS/MS approach could be confirmed in this data set, even including nine (ECE1, NPTN,

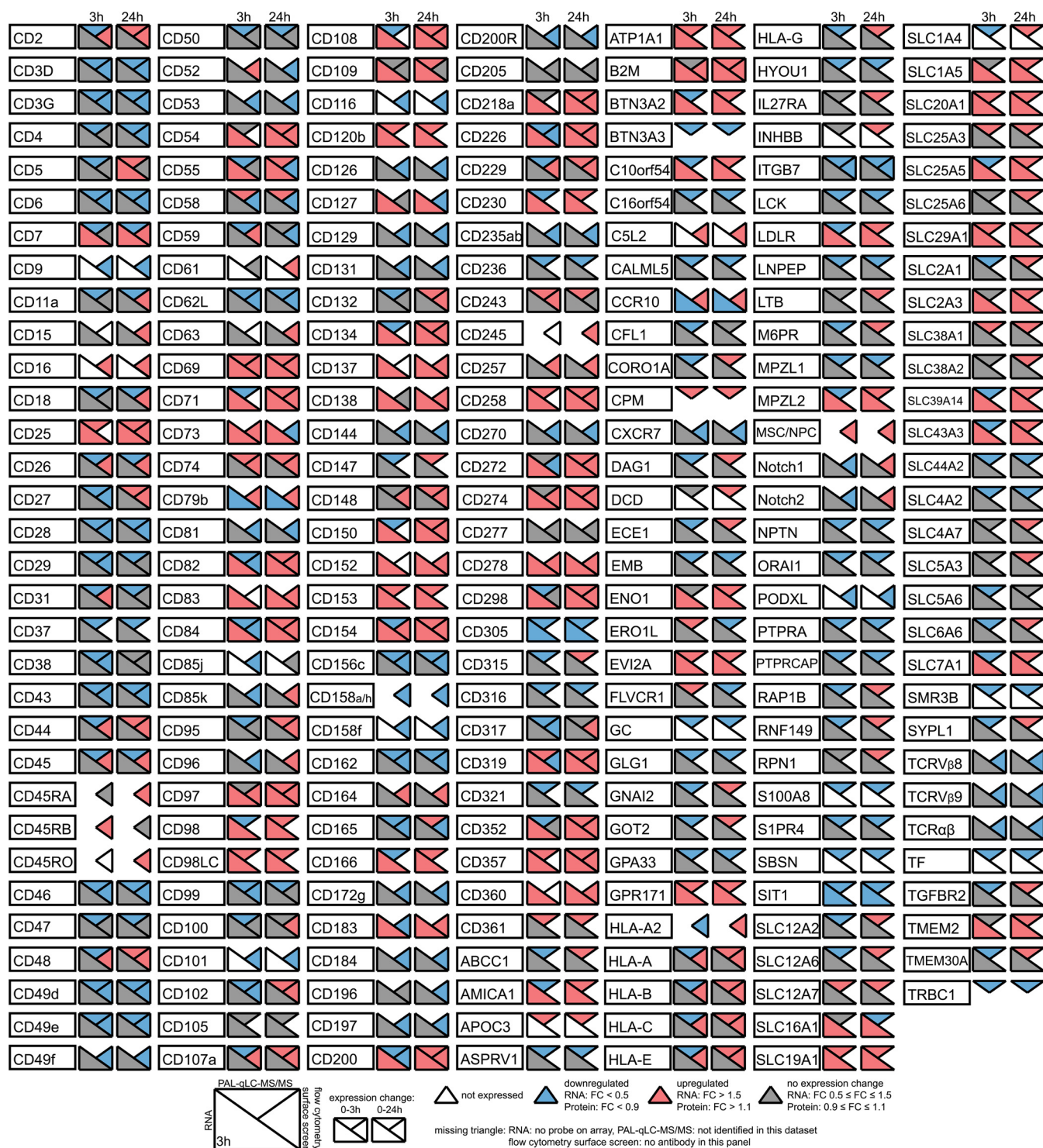


FIG. 5. The human naive CD4⁺ T-cell surface atlas. Summary of all via PAL-qLC-MS/MS and flow cytometry identified cell surface proteins. Trends of quantitative expression changes after αCD3/αCD28 stimulation of proteins (0–3 h and 0–24 h) and their respective genes (0–3 h) are given*. (*if multiple isoforms per protein or multiple probes per gene were measured, only the highest value was included in the analysis.)

RNF149, SLC1A4, SLC4A7, SLC5A3, SLC7A1, SYPL1, TMEM2) of the 24 markers, which were not linked to T cells so far. Thus, the next step will be to focus on these interesting

candidates and to prove their status as novel surface proteins. Furthermore, functional studies will be required to elucidate their potential functional relevance within T-cell biology.

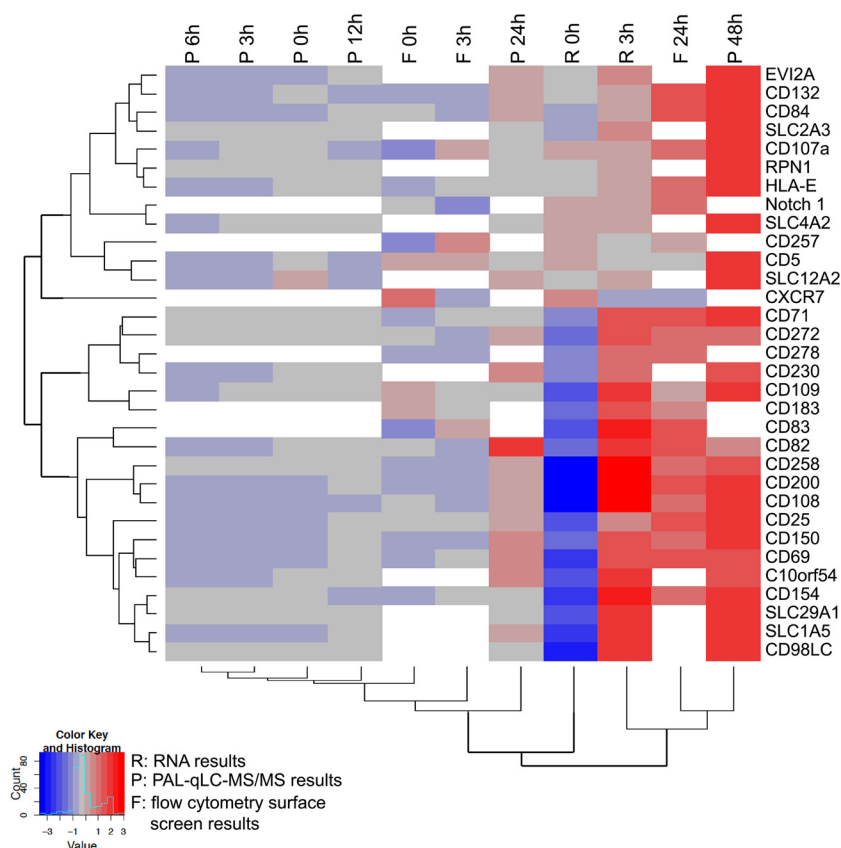


FIG. 6. **Heat map representing proteins significantly regulated in the combined omics approaches.** Proteins identified on RNA as well as on protein level (PAL-qLC-MS/MS and/or flow cytometry), showing a significant differential expression regulation in at least one of the technologies and one stimulation time point ($FC > 1$, $p < 0.01$), are listed ($n = 32$). Expression levels are represented in a color scale from blue (low expressed) to red (highly expressed).

Unsupervised clustering of the whole mass-spectrometry-based data set grouped the identified proteins into three distinct dynamic expression profiles, each of them characterized by specific Gene Ontology terms. Interestingly, 38% (cluster 1 and 2) of the proteins showed a strong down-regulation within the first hours of activation. In general, the expression of cell surface proteins is subjected to a fast turnover, which enables those proteins to respond to any kind of stimulation. Markers contained in cluster 1 and 2 such as Lck, CD3, CD4, CD28, and CD44 are present in the immunological synapse and are known for their early expression change during stimulation. Others, such as SEMA7A or CD274 (Cluster 3) are late activation markers and play mainly a role as negative regulators, which lead to the termination of T-cell responses. In line with this, Naramura *et al.* (41) showed that T-cell activation is negatively regulated by clearance/internalization of the engaged TCR from the cell surface, a process that seems to be essential for the termination of the TCR signal.

In order to expand and endorse the data set of surface proteins generated by PAL-qLC-MS/MS with an independent technique and to detect further surface markers, an antibody-based flow cytometry approach was performed. Hereby we

screened the expression of 332 surface antigens with monoclonal antibodies and could verify 67 proteins detected by PAL-qLC-MS/MS, whereas further 56 additional surface proteins were solely identified by flow cytometry.

Metcalfe *et al.* (42) recently performed a different mass-spectrometry based approach by focusing on surface proteins of murine leukocytes that contain redox labile disulfide bonds. Redox conditions change during immune activation and thus mild reducing conditions, comparable with those expected during activation, were applied to a murine 2B4 T-cell hybridoma cell line. By doing so, a large set of 87 membrane proteins containing labile disulfide bonds was identified, of which half of the proteins were confirmed by our human proteomic data set.

Of note, our proteomic surface approach is limited to the identification of N-glycoproteins by the PAL-qLC-MS/MS and restricted to available antibodies in the case of the flow cytometry approach. Thus, to overcome this drawback and to estimate the quantity of surface proteins, which might be potentially expressed on the surface of human naive CD4⁺ T cells, we continued with the transcriptomic approach. In particular, we identified genes for which the subcellular localization of their corresponding proteins was annotated in Uni-

ProtKB as being on the cell surface or in the plasma membrane. If no experimentally verified annotation could be obtained from the UniProtKB, we used predicted subcellular localizations from the recently published LocTree3 (accuracy 80%) further supported by predicted transmembrane helices (1 or more for plasma membrane proteins). We are aware, that the selection criteria for cell surface protein annotation are more strict in the transcriptional data set than in the proteomic data set. However, because experimental evidence for the subcellular localization of the protein data is provided by the PAL-qLC-MS/MS and/or the flow cytometry approach, we strengthened the criteria in the transcriptomic data set to reduce false positive results.

The transcriptomic analysis revealed 927 transcripts, which fulfilled the selection criteria. Recently, da Cunha *et al.* (43) generated an overall catalog of cell surface proteins based on the human genome sequence (National Center for Biotechnology Information build 36.1). By using bioinformatics tools a set of more than 3700 genes, believed to encode cell surface proteins of human cells, were identified. This number is in line with other estimates assuming that around 26% of all human genes code for proteins that are located at the cell surface (44). The reduced number of identified surface proteins in our transcriptomic data set compared with this study could be explained by the fact that (1) we focused specifically on the subset of CD4⁺ naive T cells and (2) we used a much stricter selection approach as described above.

Out of the 927 selected transcripts we identified in our bioinformatics analysis, 141 have to be emphasized, because those revealed a significant differential expression after 3 h of α CD3/ α CD28 activation. The top 10 genes with the highest expression changes are *SEMA7A*, *FASLG*, *CLIC4*, *FFAR3*, *CD200*, *TMEM88*, *TNF*, *SLC7A5*, *TNFSF14*, and *CD69*. For most of them, their role in T-cell activation and differentiation is well known. However others such as *FFAR3* (the free fatty acid receptor 3), which was recently described in a murine study as a crucial sensor on dendritic cell precursors for the transport of short-chain fatty acids (45), have to be further investigated on the protein level in the context of naive CD4⁺ T cells.

A comparison between the transcriptomic and the proteomic data sets showed that 53% of the proteins identified by mass-spectrometry and flow cytometry would have been identified as cell surface proteins via the bioinformatics approach alone. An even more extensive overlap of 83% between the transcriptional and the proteomic approach could be reached if the more general annotation “membrane” (subcellular localization not further classified) and the putative cell surface protein data set are included in the analysis. The putative data set allows additional subcellular annotations given by the UniProtKB such as “lipid- anchor” or “peripheral” (side of the plasma membrane to which they are attached is not specified) to be considered in the analysis and presents 248 additional genes, potentially encoding cell surface pro-

teins. On the one hand, this comparison still illustrates the limitations obtained by trusting only transcriptional approaches supported by bioinformatics analysis. Forty-seven percent of the identified proteins are not identified as cell surface proteins by the strict transcriptional data set, underscoring the essential need of proteomic approaches. On the other hand the bioinformatics approach used for the identification of cell surface proteins (including the putative data set) presents a great assembly of further genes ($n = 1074$) potentially coding for cell surface proteins, which have to be experimentally validated on a proteomic level on the surface of human naive CD4⁺ T cells. The fact that these surface membrane protein coding genes, which were identified by the bioinformatics analysis, were not detected by the two proteomic approaches might be explained by (1) not all genes present on the RNA level are translated into proteins, (2) proteins are subjected to post-translational modifications, translocation, and degradation, and (3) not all of the genes code for proteins containing a glycosylation site. A potential N-glycosylation site was proposed for 69% of the genes coding for cell surface proteins by the UniProtKB database (supplemental Table S3). Thus, one might speculate that we are missing around 30% of the proteins expressed on the surface of naive CD4⁺ T cells using the PAL approach. However, because we are aware of this limitation, we added the flow cytometry approach and the transcriptome data set in order to expand the cell surface atlas and to introduce a promising collection of further potential surface proteins to the interested readership.

In general, we observed only a weak correlation between the expression changes of transcripts and the corresponding translated proteins, which is already shown in various previous studies (46, 47). Nonetheless, by combining the transcriptional and the proteomic approaches (PAL-qLC-MS/MS and/or flow cytometry), 32 out of 229 cell surface proteins showed a significant expression change in at least one technical approach for a minimum of one stimulation time point. As expected, the majority of these proteins like e.g. CD69 (T-cell activation marker) and CD108 (known to modulate inflammation and T-cell mediated immune response) are well characterized in the context of T-cell biology. Within this data set, 21 cell surface proteins are already annotated as CD molecules. Six proteins belong to the solute carrier family and one of them, SLC7A5 (CD98LC), was recently described as the central large neutral amino acid transporter in a murine study (48), indispensable for the effector differentiation and clonal expansion of T cells upon antigenic TCR stimulation. In the current human study, this surface transporter protein was also among the top significantly regulated hits on RNA as well as on protein level. SLC7A5 showed a strong up-regulation upon the general α CD3/ α CD28 TCR stimulus like it was previously shown for CD4⁺/CD25⁻ T cells (49). Further members of the SLC family, (SLC4A2-Anion exchanger 2, SLC2A3-GLUT-3, SLC29A1-equilibrative nucleoside transporter 1,

SLC1A5-neutral amino acid transporter B, SLC12A2-Na⁺/K⁺/Cl⁻ cotransporter) were detected, but all of them are missing the affiliation to the cluster of differentiation (CD). Although their general role as transporters is described, it needs to be further validated if they play a similar important role in T-cell signaling as SLC7A5. Furthermore, EVI2A (ecotropic viral integration site 2A) was identified in our analysis as an interesting naive CD4⁺ T-cell marker. Together with *EVI2B*, it lies in a long intron of *NF1* (neurofibromin 1), which is known to be completely deleted in some patients with neurofibromatosis (50, 51). It is also known that the EVI2 locus (murine homolog of human EVI2A) is an integration site for the chronic Murine Leukemia Virus, possibly leading to myeloid leukemias upon infection and integration of the virus at this gene locus (52). This protein is thought to be arranged in complexes as homodimer or with other proteins located in the plasma membrane as a heterodimer to form a cell surface receptor (information given by UniProtKB). Its neighbor gene, *EVI2B*, which was also detected on transcriptional and protein level in our cell surface approach, but did not exhibit a significant regulation, was accepted in the last Human Leukocyte Differentiation Antigens (HLDA) Workshop in 2010 (53) as a new member of the CD family and nomenclatured as CD361. *EVI2B* was assigned as a new B cell marker, but its expression on T cells, monocytes, granulocytes and NK cells was additionally proven, leading to the assumption that this protein might have a general role on cells of the immune system. However, until now, no functional characterization of either *EVI2B* nor *EVI2A* has occurred leaving their function unknown. Thus, future studies are essential to identify their potential function in the context of T-cell activation.

The high translational potential of cell surface protein focusing technologies was already shown by data from Mirkowska *et al.* (54) by the analysis of the surface composition of xenografts from 19 B-cell precursor acute lymphoblastic leukemia patients leading to the identification of new diagnostic leukemia-associated candidate markers. In a second study, the surface proteome of myeloid leukemia cells was identified to target antigens, which may be used for the development of therapeutic antibodies (55). Cell surface proteins are easy accessible targets for drugs, especially for antibodies. Seven proteins of our data set are already described as targets for approved therapeutic antibodies: therapeutic antibodies against CD52 and CD152 are applied in cancer therapy, whereas CD3, CD25, CD11a, CD49d, and CD126 are targeted in the context of autoimmune diseases (www.antibodysociety.org, www.immunologylink.com). Furthermore, the high translational potential of the presented cell surface atlas becomes obvious by subjecting the identified proteins ($n = 229$) to a DrugBank database search (56). In total, the current study identified 60 proteins (Table I), which are targets for approved ($n = 51$) or experimentally validated/currently under investigation drugs ($n = 9$).

Taken together, the combination of the transcriptomic as well as the proteomic data sets enabled us to generate an extensive cell surface atlas of human naive and activated CD4⁺ T cells on an experimental level and introduces a promising collection of further candidate proteins. Furthermore, the cell surface atlas can serve as a reference book for the mapping of cell surface proteins on human naive and activated CD4⁺ T cells. It presents known markers and proposes potential new drug targets such as EVI2A usable for the modification of T-cell activation.

Acknowledgments—We thank B.Grothe (Division of Neurobiology, Department Biology II, Ludwig-Maximilians-Universität München, Germany) for support of the studies.

*E.K. is supported by NIH grant GM095315. T.G. is supported by Alexander von Humboldt Foundation through German Federal Ministry for Education and Research; Ernst Ludwig Ehrlich Studienwerk. M.S. is supported by the German Research Foundation and the Else Kröner-Fresenius Stiftung. B.K. is supported by the German Research Foundation (SPP1395/InKoMBio Busch 900/6-1).

§ This article contains supplemental Figs. S1 to S6 and Tables S1 to S7.

To whom correspondence should be addressed: Center of Allergy and Environment (ZAUM), Technische Universität und Helmholtz Zentrum München, Biedersteiner Straße 29, 80802 München, Germany. Tel.: +49 89 4140-3464; Fax: +49 89 4140-3453; E-mail: Kathrin.Suttner@lrz.tum.de.

The MS proteomics data in this paper have been deposited in the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository data set identifier PXD001432. The RNA array data set was deposited in NCBI's Gene Expression Omnibus and is accessible through GEO Series accession number GSE61983. This work was supported by iMed - the Helmholtz Initiative on Personalized Medicine. This work is part of the PhD thesis of A.G. and is supported by HELENA, Helmholtz Graduate School Environmental Health, Helmholtz Zentrum München, Neuherberg, Germany.

REFERENCES

1. Zhu, J., Yamane, H., and Paul, W. E. (2010) Differentiation of effector CD4 T cell populations. *Annu. Rev. Immunol.* **28**, 445–489
2. Eyerich, S., Onken, A. T., Weidinger, S., Franke, A., Nasorri, F., Pennino, D., Grosber, M., Pfab, F., Schmidt-Weber, C. B., Mempel, M., Hein, R., Ring, J., Cavani, A., and Eyerich, K. (2011) Mutual antagonism of T cells causing psoriasis and atopic eczema. *N. Engl. J. Med.* **365**, 231–238
3. Eyerich, K., Pennino, D., Scarponi, C., Foerster, S., Nasorri, F., Behrendt, H., Ring, J., Traidl-Hoffmann, C., Albanesi, C., and Cavani, A. (2009) IL-17 in atopic eczema: linking allergen-specific adaptive and microbial-triggered innate immune response. *J. Allergy Clin. Immunol.* **123**, 59–66 e54
4. Wei, G., Abraham, B. J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D. L., Tang, Q., Paul, W. E., Zhu, J., and Zhao, K. (2011) Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* **35**, 299–311
5. Yosef, N., Shalek, A. K., Gaublot, J. T., Jin, H., Lee, Y., Awasthi, A., Wu, C., Karwacz, K., Xiao, S., Jorgoli, M., Gennert, D., Satija, R., Shakya, A., Lu, D. Y., Trombetta, J. J., Pillai, M. R., Ratcliffe, P. J., Coleman, M. L., Bix, M., Tantin, D., Park, H., Kuchroo, V. K., and Regev, A. (2013) Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496**, 461–468
6. Hawkins, R. D., Larjo, A., Tripathi, S. K., Wagner, U., Luu, Y., Lonnberg, T., Raghav, S. K., Lee, L. K., Lund, R., Ren, B., Lahdesmaki, H., and Lahesmaa, R. (2013) 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
7. Meissner, F., and Mann, M. (2014) Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology. *Nat. Immunol.* **15**, 112–117
 8. Lonnberg, T., Chen, Z., and Laheesmaa, R. (2013) From a gene-centric to whole-proteome view of differentiation of T helper cell subsets. *Brief Funct. Genomics* **12**, 471–482
 9. Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587
 10. Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaer-kady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T. C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., and Pandey, A. (2014) A draft map of the human proteome. *Nature* **509**, 575–581
 11. Nyman, T. A., Rosengren, A., Syrakki, S., Pellinen, T. P., Rautajoki, K., and Laheesmaa, R. (2001) A proteome database of human primary T helper cells. *Electrophoresis* **22**, 4375–4382
 12. Kobayashi, M., Katagiri, T., Kosako, H., Iida, N., and Hattori, S. (2007) Global analysis of dynamic changes in lipid raft proteins during T-cell activation. *Electrophoresis* **28**, 2035–2043
 13. de Wet, B., Zech, T., Salek, M., Acuto, O., and Harder, T. (2011) Proteomic characterization of plasma membrane-proximal T cell activation responses. *J. Biol. Chem.* **286**, 4072–4080
 14. Bini, L., Pacini, S., Liberatori, S., Valensin, S., Pellegrini, M., Raggiaschi, R., Pallini, V., and Baldari, C. T. (2003) Extensive temporally regulated reorganization of the lipid raft proteome following T-cell antigen receptor triggering. *Biochem. J.* **369**, 301–309
 15. Moulder, R., Lonnberg, T., Elo, L. L., Filen, J. J., Rainio, E., Corthals, G., Oresic, M., Nyman, T. A., Aittokallio, T., and Laheesmaa, R. (2010) 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
 16. Ruperez, P., Gago-Martinez, A., Burlingame, A. L., and Osés-Prieto, J. A. (2012) 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
 17. Loyet, K. M., Ouyang, W., Eaton, D. L., and Stults, J. T. (2005) Proteomic profiling of surface proteins on Th1 and Th2 cells. *J. Proteome Res.* **4**, 400–409
 18. Zeng, Y., Ramya, T. N., Dirksen, A., Dawson, P. E., and Paulson, J. C. (2009) High-efficiency labeling of sialylated glycoproteins on living cells. *Nat. Methods* **6**, 207–209
 19. Hauck, S. M., Dietter, J., Kramer, R. L., Hofmaier, F., Zipplies, J. K., Amann, B., Feuchtinger, A., Deeg, C. A., and Ueffing, M. (2010) Deciphering membrane-associated molecular processes in target tissue of autoimmune uveitis by label-free quantitative mass spectrometry. *Mol. Cell. Proteomics* **9**, 2292–2305
 20. Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., and Watts, J. D. (2009) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* **27**, 378–386
 21. Gundry, R. L., Riordon, D. R., Tarasova, Y., Chuppa, S., Bhattacharya, S., Juhász, O., Wiedemeier, O., Milanovich, S., Noto, F. K., Tchernyshyov, I., Raginski, K., Bausch-Fluck, D., Tae, H. J., Marshall, S., Duncan, S. A., Wollscheid, B., Wersto, R. P., Rao, S., Van Eyk, J. E., and Boheler, K. R. (2012) A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol. Cell. Proteomics* **11**, 303–316
 22. Niehage, C., Steenblock, C., Pursche, T., Bornhauser, M., Corbeil, D., and Hoflack, B. (2011) The cell surface proteome of human mesenchymal stromal cells. *PLoS One* **6**, e20399
 23. Moest, H., Frei, A. P., Bhattacharya, I., Geiger, M., Wollscheid, B., and Wolfrum, C. (2013) Malfunctioning of adipocytes in obesity is linked to quantitative surfaceome changes. *Biochim. Biophys. Acta* **1831**, 1208–1216
 24. Deeb, S. J., Cox, J., Schmidt-Supprian, M., and Mann, M. (2014) N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Mol. Cell. Proteomics* **13**, 240–251
 25. Weekes, M. P., Antrobus, R., Lill, J. R., Duncan, L. M., Hor, S., and Lehner, P. J. (2010) Comparative analysis of techniques to purify plasma membrane proteins. *J. Biomol. Tech.* **21**, 108–115
 26. Merl, J., Ueffing, M., Hauck, S. M., and von Toerne, C. (2012) Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal Muller cells. *Proteomics* **12**, 1902–1911
 27. von Toerne, C., Menzler, J., Ly, A., Senninger, N., Ueffing, M., and Hauck, S. M. (2014) Identification of a novel neurotrophic factor from primary retinal Muller cells using stable isotope labeling by amino acids in cell culture (SILAC). *Mol. Cell. Proteomics* **13**, 2371–2381
 28. Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dienes, J. A., Sun, Z., Farrah, T., Bandeira, N., Binz, P. A., Xenarios, I., Eisenacher, M., Mayer, G., Gatto, L., Campos, A., Chalkley, R. J., Kraus, H. J., Albar, J. P., Martinez-Bartolome, S., Apweiler, R., Omenn, G. S., Martens, L., Jones, A. R., and Hermjakob, H. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226
 29. Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Giron, C. G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kahari, A. K., Keenan, S., Kulesha, E., Martin, F. J., Maurel, T., McLaren, W. M., Murphy, D. N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H. S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S. J., Vullo, A., Wilder, S. P., Wilson, M., Zadissa, A., Aken, B. L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T. J., Kinsella, R., Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D. R., and Searle, S. M. (2014) Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755
 30. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42**, D191–D198
 31. Kumar, L., and M. E. F. (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2**, 5–7
 32. Rigbolt, K. T., Vanselow, J. T., and Blagoev, B. (2011) GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Mol. Cell. Proteomics* **10**, O110 007450
 33. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29
 34. Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., and Sherlock, G. (2004) 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
 35. Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011) REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLoS One* **6**, e21800
 36. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210
 37. Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C. M., Hart, J., Landrum, M. J., McGarvey, K. M., Murphy, M. R., O'Leary, N. A., Pujar, S., Rajput, B., Rangwala, S. H., Riddick, L. D., Shkeda, A., Sun, H., Tamez, P., Tully, R. E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D. R., Murphy, T. D., and Ostell, J. M. (2014) RefSeq: an update on mammalian

- reference sequences. *Nucleic Acids Res.* **42**, D756–D763
38. Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., Altermann, U., Angerer, P., Ansorge, S., Balasz, K., Bernhofer, M., Betz, A., Cizmadija, L., Do, K. T., Gerke, J., Greil, R., Joerdens, V., Hastreiter, M., Hembach, K., Herzog, M., Kalemanov, M., Kluge, M., Meier, A., Nasir, H., Neumaier, U., Prade, V., Reeb, J., Sorokoumov, A., Troshani, I., Vorberg, S., Waldruff, S., Zierer, J., Nielsen, H., and Rost, B. (2014) LocTree3 prediction of localization. *Nucleic Acids Res.* **42**, W350–W355
39. Kall, L., Krogh, A., and Sonnhammer, E. L. (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **1**, i251–i257
40. Bausch-Fluck, D., Hofmann, A., Bock, T., Frei, A. P., Cerciello, F., Jacobs, A., Moest, H., Omasits, U., Gundry, R. L., Yoon, C., Schiess, R., Schmidt, A., Mirkowska, P., Hartlova, A., Van Eyk, J. E., Bourquin, J. P., Aebersold, R., Boheler, K. R., Zandstra, P., and Wollscheid, B. (2015) A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314
41. Naramura, M., Jang, I. K., Kole, H., Huang, F., Haines, D., and Gu, H. (2002) c-Cbl and Cbl-b regulate T cell responsiveness by promoting ligand-induced TCR down-modulation. *Nat. Immunol.* **3**, 1192–1199
42. Metcalfe, C., Cresswell, P., Ciaccia, L., Thomas, B., and Barclay, A. N. (2011) Labile disulfide bonds are common at the leucocyte cell surface. *Open Biol.* **1**, 110010
43. da Cunha, J. P., Galante, P. A., de Souza, J. E., de Souza, R. F., Carvalho, P. M., Ohara, D. T., Moura, R. P., Oba-Shinja, S. M., Marie, S. K., Silva, W. A., Jr., Perez, R. O., Stransky, B., Pieprzyk, M., Moore, J., Caballero, O., Gama-Rodrigues, J., Habr-Gama, A., Kuo, W. P., Simpson, A. J., Camargo, A. A., Old, L. J., and de Souza, S. J. (2009) Bioinformatics construction of the human cell surfaceome. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16752–16757
44. Fagerberg, L., Jonasson, K., von Heijne, G., Uhlen, M., and Berglund, L. (2010) Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149
45. Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L., and Marsland, B. J. (2014) Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* **20**, 159–166
46. Rugg-Gunn, P. J., Cox, B. J., Lanner, F., Sharma, P., Ignatchenko, V., McDonald, A. C., Garner, J., Gramolini, A. O., Rossant, J., and Kislinger, T. (2012) Cell-surface proteomics identifies lineage-specific markers of embryo-derived stem cells. *Dev. Cell* **22**, 887–901
47. Boheler, K. R., Bhattacharya, S., Kropp, E. M., Chuppa, S., Riordon, D. R., Bausch-Fluck, D., Burridge, P. W., Wu, J. C., Wersto, R. P., Chan, G. C., Rao, S., Wollscheid, B., and Gundry, R. L. (2014) A human pluripotent stem cell surface N-glycoproteome resource reveals markers, extracellular epitopes, and drug targets. *Stem Cell Rep.* **3**, 185–203
48. Sinclair, L. V., Rolf, J., Emslie, E., Shi, Y. B., Taylor, P. M., and Cantrell, D. A. (2013) Control of amino-acid transport by antigen receptors coordinates the metabolic reprogramming essential for T cell differentiation. *Nat. Immunol.* **14**, 500–508
49. Hayashi, K., Jutabha, P., Endou, H., Sagara, H., and Anzai, N. (2013) LAT1 is a critical transporter of essential amino acids for immune reactions in activated human T cells. *J. Immunol.* **191**, 4080–4085
50. Kayes, L. M., Burke, W., Riccardi, V. M., Bennett, R., Ehrlich, P., Rubenstein, A., and Stephens, K. (1994) Deletions spanning the neurofibromatosis 1 gene: identification and phenotype of five patients. *Am. J. Hum. Genet.* **54**, 424–436
51. Jenne, D. E., Tinschert, S., Stegmann, E., Reimann, H., Nurnberg, P., Horn, D., Naumann, I., Buske, A., and Thiel, G. (2000) A common set of at least 11 functional genes is lost in the majority of NF1 patients with gross deletions. *Genomics* **66**, 93–97
52. Buchberg, A. M., Bedigian, H. G., Jenkins, N. A., and Copeland, N. G. (1990) Evi-2, a common integration site involved in murine myeloid leukemogenesis. *Mol. Cell. Biol.* **10**, 4658–4666
53. Matesanz-Isabel, J., Sintes, J., Llinas, L., de Salort, J., Lazaro, A., and Engel, P. (2011) New B-cell CD molecules. *Immunol. Lett.* **134**, 104–112
54. Mirkowska, P., Hofmann, A., Sedek, L., Slamova, L., Mejstrikova, E., Szczepanski, T., Schmitz, M., Cario, G., Stanulla, M., Schrappe, M., van der Velden, V. H., Bornhauser, B. C., Wollscheid, B., and Bourquin, J. P. (2013) Leukemia surfaceome analysis reveals new disease-associated features. *Blood* **121**, e149–159
55. Strassberger, V., Gutbrodt, K. L., Krall, N., Roesli, C., Takizawa, H., Manz, M. G., Fugmann, T., and Neri, D. (2014) A comprehensive surface proteome analysis of myeloid leukemia cell lines for therapeutic antibody development. *J. Proteomics* **99**, 138–151
56. Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z. T., Han, B., Zhou, Y., and Wishart, D. S. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097