ISCI, Volume *11*

Supplemental Information

Quantitative Proteomics Reveals the Dynamic Protein Landscape during Initiation of Human Th17 Cell Polarization

Subhash K. Tripathi, Tommi Välikangas, Ankitha Shetty, Mohd Moin Khan, Robert Moulder, Santosh D. Bhosale, Elina Komsi, Verna Salo, Rafael Sales De Albuquerque, Omid Rasool, Sanjeev Galande, Laura L. Elo, and Riitta Lahesmaa

Supplemental Information

Title: Quantitative proteomics reveals the dynamic protein landscape during initiation of human Th17 cell polarization

One Sentence Summary: Proteome analysis of early human Th17 cell differentiation

Authors: Subhash K. Tripathi[1±], Tommi Välikangas[1,2±], Ankitha Shetty[1,4±], Mohd Moin Khan[1,3], Santosh D Bhosale[1,3], Robert Moulder[1], Elina Komsi[1], Verna Salo[1,3], Rafael Sales De Albuquerque[1], Omid Rasool[1], Sanjeev Galande[4], Laura L. Elo[*1±] and Riitta Lahesmaa[*1±]

Supplementary figures and legends

Supplementary Figure 1 (related to figure 1)

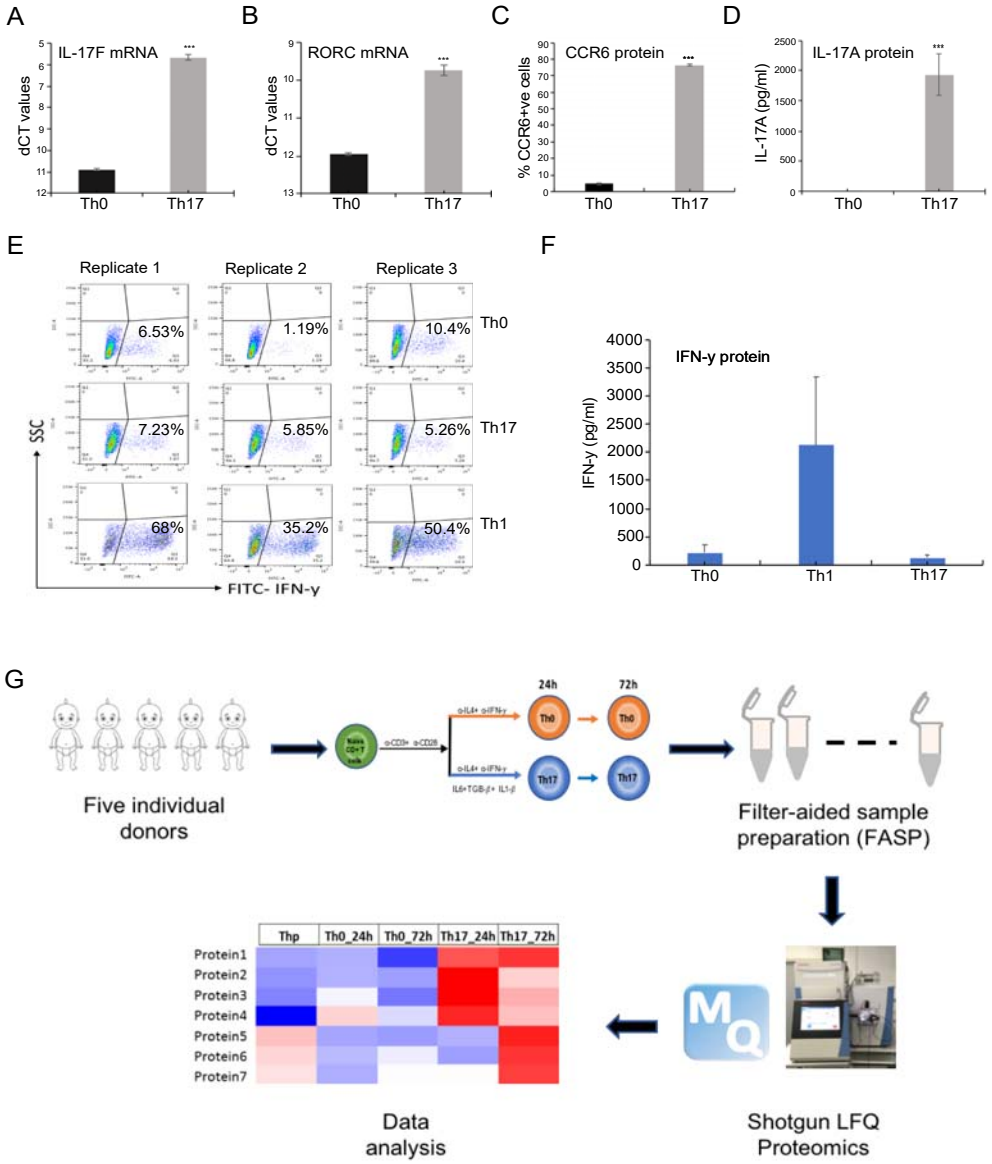
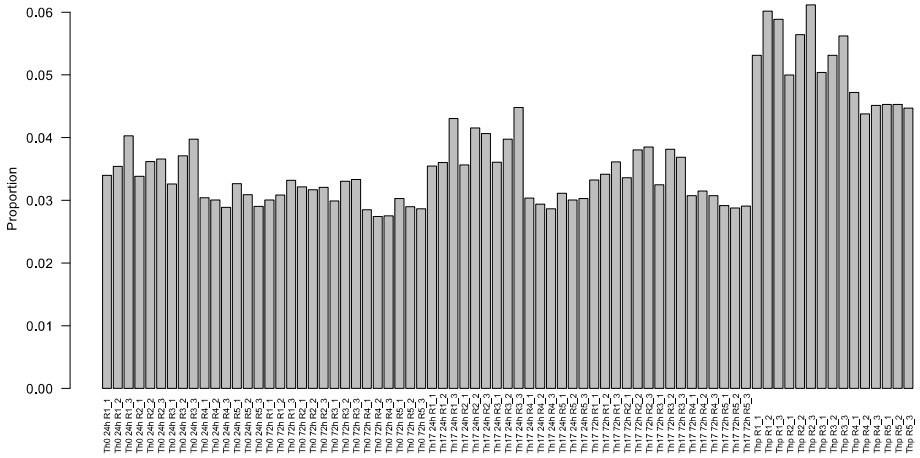


Figure S1. The expression of Th17 markers in Th17 cultures and a schematic representation of the study setting. (A) IL17F mRNA expression, (B) RORC mRNA expression, (C) CCR6 protein expression, and (D) IL17A protein secretion from the cultures at 72h post initiation of Th17 polarization. The data is from five individual donors. Error bars show the standard deviation. P-values were calculated using the two-tailed Student's t-test. (E) The expression of IFN-γ was measured by intracellular staining using flow cytometry and (F) the secretion of IFN-γ with Luminex analysis from three biological replicates. (G) A schematic representation of the study. Error bars represent standard deviation across biological replicates (n = 3-5; P <0.001(**) for Th0 vs Th17; paired t-test analysis).

Supplementary Figure 2: Related to Figure 2.

A



B

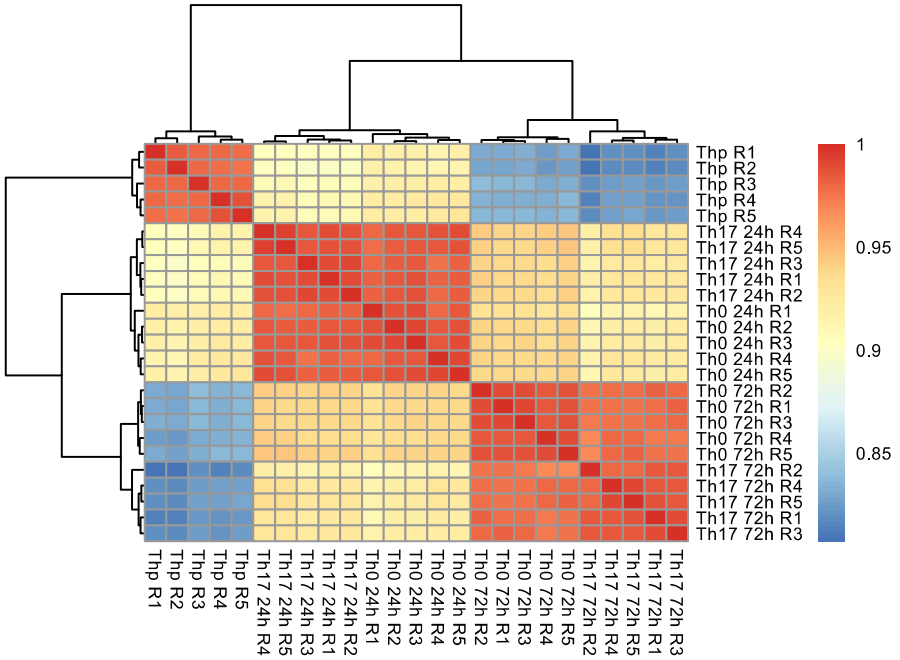


Figure S2. Quality of the proteomics data. (A) The proportion of missing values in each sample. (B) The similarity and clustering of the samples. Pearson correlation coefficient was used as a similarity measure and the hierarchical clustering with complete linkage was used for the clustering.

Supplementary Figure 3: Related to Figure 3.

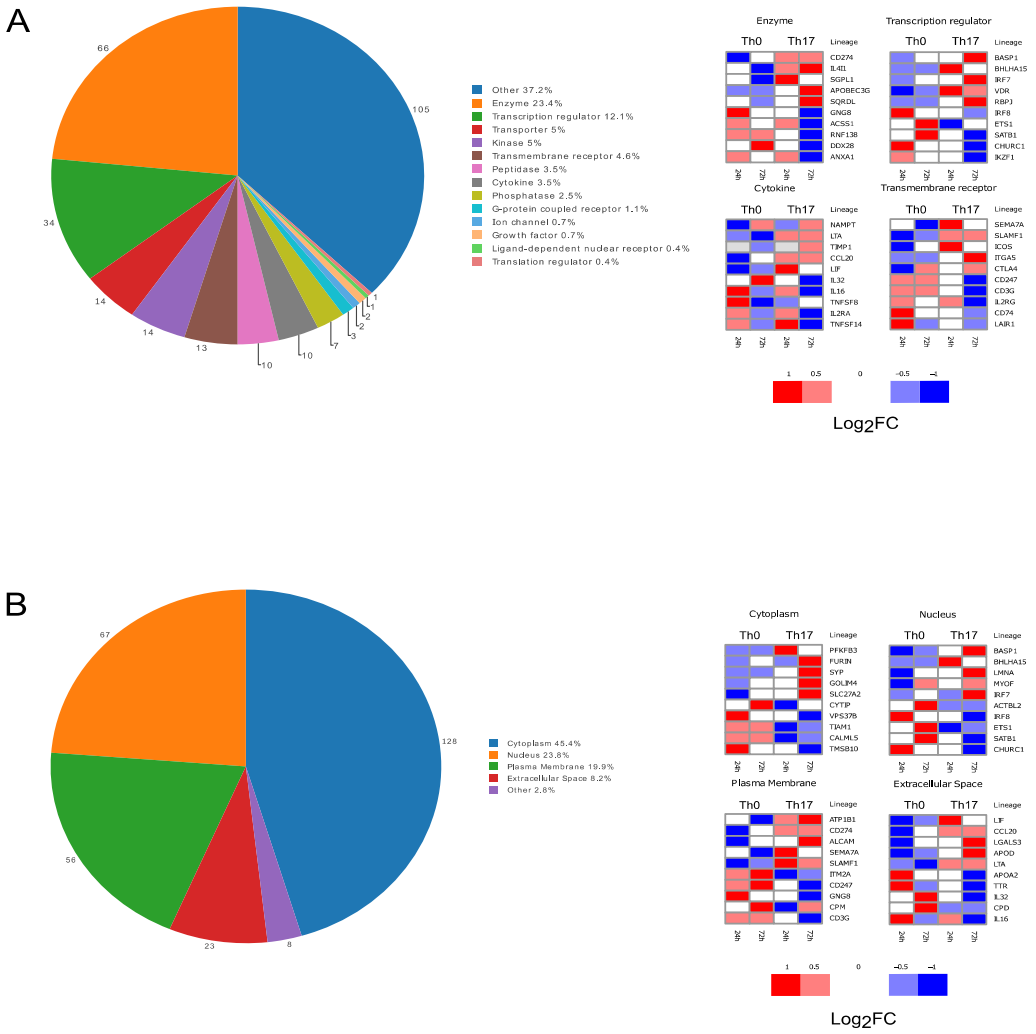


Figure S3. The functional types, and cellular locations among the differentially expressed proteins between Th17 and Th0. (A) Proportions and counts of the IPA functional types associated with the differentially expressed proteins over both time-points and the standardized averaged expression of the most up- and down-regulated proteins based on average logarithmic fold changeover both time-points for the selected types. (B) Proportions and counts of the IPA cellular locations associated with the DE proteins over both timepoints and the standardized averaged expression of the most up- and down-regulated proteins based on average logarithmic fold change over both timepoints for the selected locations.

Supplementary Figure 4 (related to figure 4)

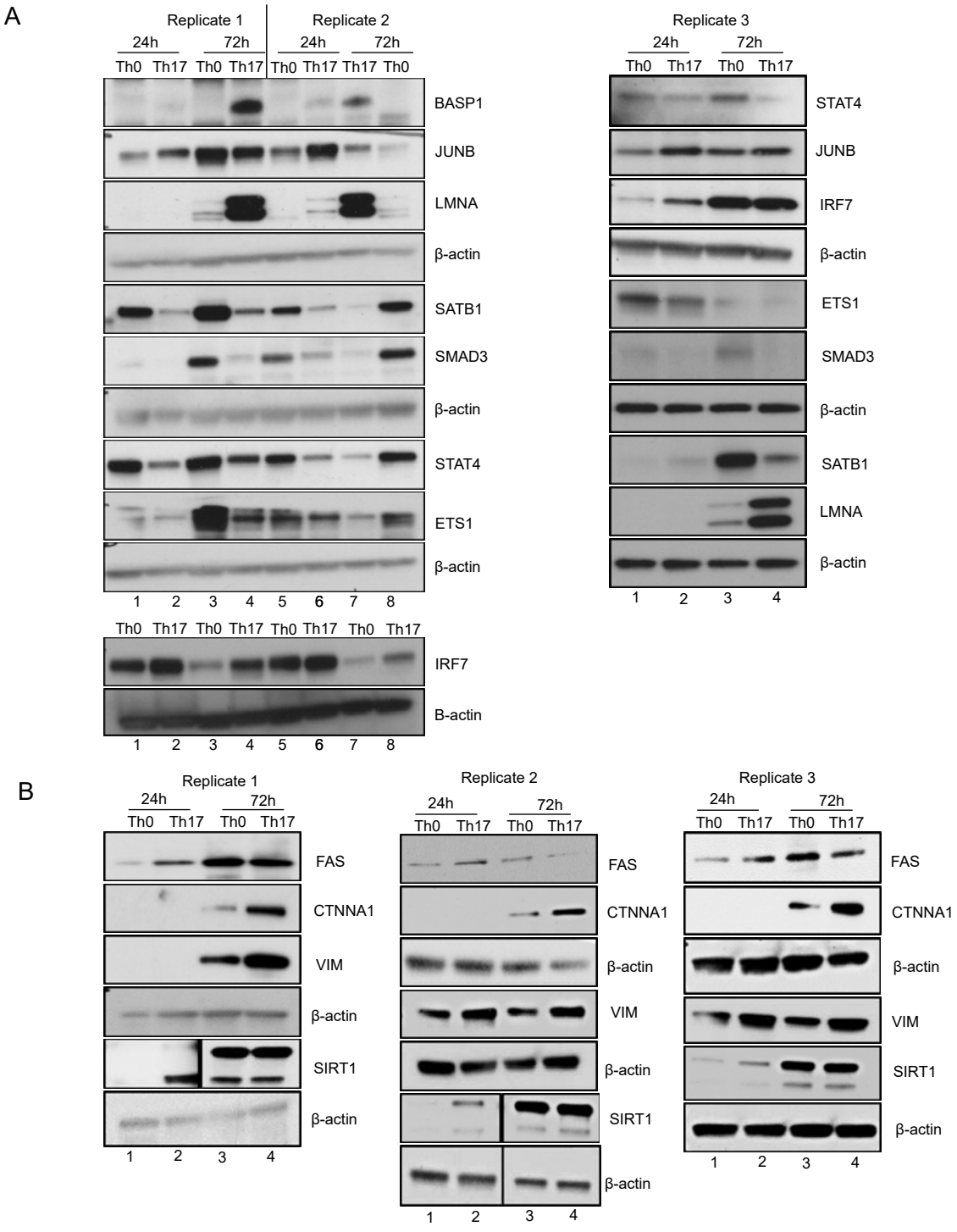
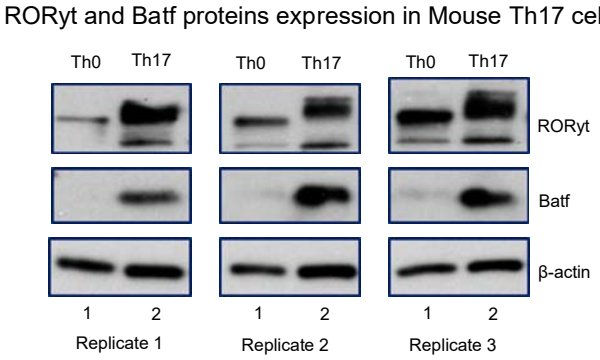


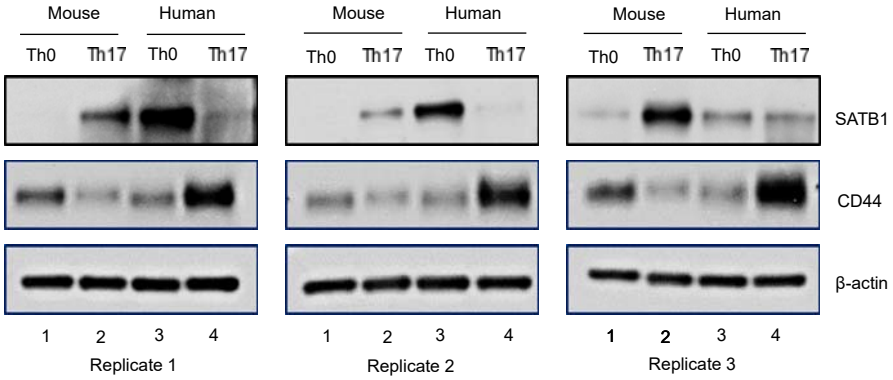
Figure S4. Validation of the Mass-spectrometry (MS) identified proteins using Immunoblotting. (A and B), Immunoblot validations of MS identified up- or downregulated proteins with known and unknown Th17-related function. Blots show the protein extracts from the Th0 and Th17 cells at 24h and 72h.

Supplementary Figure 5 (related to Figure 6 and 7)

A Related to figure 6



B Related to figure 6



C Related to figure 7

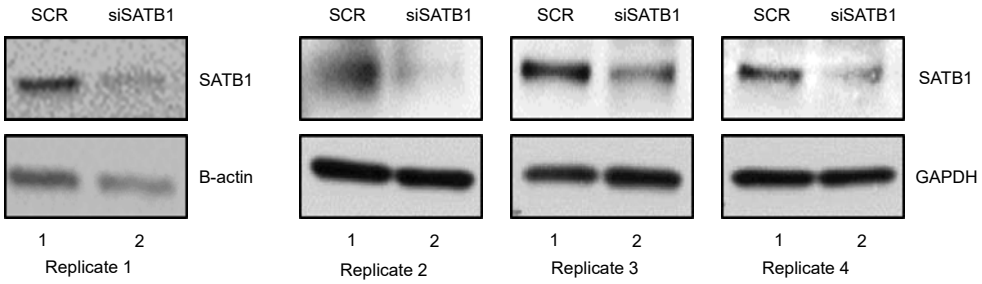


Figure S5. Validation of opposite expression of SATB1 and CD44 in human and mouse during Th17 polarization and SATB1 expression upon SATB1 RNAi silencing. (A) Immunoblot analysis of Th17 marker transcription factors, RORyt and BATF expression during mouse Th17 cell polarization, (B) Immunoblot analysis of SATB1 and CD44 expression in human and mouse from the Th0 and Th17 cells at 72h, and (C) Immunoblot analysis of SATB1 expression upon SATB1 knockdown from four biological replicates.

Inventory of Supplementary Table Legends

All the supplementary tables are provided as separate Excel spreadsheet and can be found in online version.

Table S1. The normalized data, missing value proportions of the samples, Pearson correlations between the samples and the differential expression results from all the comparisons. (related to Figure1 and Supplementary Figure 2)

Table S2. The up- and downregulated proteins between all cell types (Thp, Th0 and Th17) and the z-score standardized expression values of the differentially expressed proteins between Th17 and Th0. Supplementary Table information related to Figure 1.

Table S3. The GO biological process enrichment information relating to Figure 2, and the IPA type and location information relating to Supplementary Figure 3.

Table S4: The nodes, edges, clusters and enrichment information of the protein-protein interaction networks relating to Figure 4.

Table S5: The common and unique genes and proteins between the proteomics and the transcriptomics datasets at 24h, 72h, and overall and the averaged z-score standardized expression values for the common genes in proteomics and transcriptomics. Supplementary Table information related to Figure 5.

Supplementary Table 6: Supplementary Table information related to Figure 6.

Supplementary Table 7: Supplementary table information for antibodies, quantitative PCR primer and probes and SATB1 siRNA sequence.

Transparent Methods:

Primary human CD4+ T-cell isolation and Th17 cell culture

To obtain naive CD4+ T cells, human peripheral blood mononuclear cells (PBMCs) were isolated from the umbilical cord blood of healthy neonates (Turku University Central Hospital, Turku, Finland) by by the Ficoll- Paque density gradient method (Ficoll-Paque PLUS; GE Healthcare). Naive CD4+ cells were further purified using CD4+ Dynal positive selection beads (Dynal CD4 Positive Isolation Kit; Invitrogen), and after the isolation, cells from individual donors were either activated directly or pooled (in case of validation experiments) before activation. CD4+ T cells were stimulated with plate-bound αCD3 (3750 ng/6-well culture plate well; Immunotech) and soluble αCD28 (1 µg/mL; Immunotech) in a density of 2.5 × 10^6 cells/mL of X-vivo 20 serum-free medium (Lonza). The X-vivo 20 medium was supplemented with L-glutamine (2 mM, Sigma-Aldrich), and antibiotics (50 U/mL penicillin and 50 µg/mL streptomycin; Sigma-Aldrich). Th17 cell polarization was initiated with a cytokine cocktail of IL-6 (20 ng/mL; Roche), IL-1β (10 ng/mL) and TGF-β (10 ng/mL) in the presence of neutralizing anti-IFNγ (1 µg/mL) and anti-IL-4 (1 µg/mL) to block Th1 and Th2 differentiation, respectively. For the control cells (Th0), CD4+ T cells were TCR stimulated with αCD3 and αCD28 in the presence of neutralizing antibodies without differentiating cytokines and cultured in parallel. All cytokines and neutralizing antibodies used in the study were purchased from R&D Systems unless otherwise stated. All cell cultures were maintained at 37°C in a humidified atmosphere of 5% (v/v) CO_2/air.

Isolation of mouse cells and in vitro cell culture

BALB/c mice were purchased from the University of Turku animal facility. Animals were handled and housed in accordance with the University of Turku animal welfare guidelines. Cells were obtained from spleens of 8- to 10-week-old mice. Spleens were first macerated

using a cell strainer and syringe plunger to make a single cell suspension, red blood cells were removed using ACK lysis buffer (Gibco by life technology, cat# A10492-01). Cells were then isolated by positive selection using monoclonal antibodies to CD4+CD62L+ coupled with magnetic beads (MACS Miltenyi Biotec; cat# 130-106-643) using a MACS preparation MACS LS/MS column (MACS Miltenyi Biotec).

Cell cultures were performed in IMDM (Gibco) media supplemented with 5% fetal calf serum, 2 mM L-glutamine (Sigma-Aldrich), and 100 U/mL penicillin and 100 µg/mL streptomycin (Sigma-Aldrich) and 50 µM β-mercaptoethanol (Gibco). Cells were activated by plate-bound α-CD3 (1 µg/mL; BD PharMingen, cat# 553238) and soluble α-CD28 (2 µg/mL; BD PharMingen, cat# 557393) for 3 days (unless otherwise indicated) and cultured either under neutral conditions (Th0, TCR control) or Th17 differentiation conditions, which were induced by culturing the cells in the presence of TGFβ (1 ng/ml; R&D, cat# 240-B), IL6 (20 ng/ ml; R&D, cat# 406-ML), and IL-1β (10 ng/ml; R&D, cat# 201-LB). Neutralising antibodies anti-IFN-y (cat# 557530), and anti-IL-4(cat# 559062) (both at 10 µg/mL, BD PharMingen) were added to both neutral and Th17 differentiation conditions.

Mass-spectrometry sample preparation, preprocessing and preliminary analysis

Cell lysis: Proteins were extracted from the cell pellet using a lysis buffer (4% SDS, 0.1 M DTT, 0.1 M Tris-HCl, pH 7.6), heated at 95ºC for 5 min. The lysate was then sonicated at high voltage with a setting of 5 cycles for 30 seconds and 30 seconds rest between cycles. The cell debris were cleared by centrifugation at 16000g for 20 min, and a DC Protein Assay (#5000116, BioRad) was used to estimate protein amounts.

Filter Aided Sample Preparation (FASP) Method: Briefly, an aliquot corresponding to 50 µg of protein from each biological replicate corresponding to different time points (i.e., 24h and 72h) (n=5 for both Th17 and corresponding Th0 controls) were mixed with FASP urea buffer

(8 M urea in 0.1 M Tris-HCl, pH 8.5) in a 30-kDa filter tube (Millipore) to eliminate the SDS. The proteins were reduced with dithiothretiol (DTT) and alkylated with iodoacetamide in dark for 20 min. Finally, they were digested with sequencing grade modified trypsin in 1:30 (protein:protease) ratio overnight at 37ºC. The digested peptides were then acidified and desalted using Sep-Pak C18 cartridges (WAT054955 Vac 1 cc 50 mg, Waters). The desalted samples were dried using a centrifugal evaporator (Thermo Scientific) and stored at -80ºC until further LC-MS/MS analysis.

Mass spectrometry analysis: The dried peptides were reconstituted in formic acid/acetonitrile mixture, and an amount corresponding to 400 ng was analysed using Easy-nLC 1200 coupled to Q Exactive HF mass spectrometer (Thermo Scientific). The peptides were separated on 75 µm ID X 40-cm HPLC column, packed in-house with 1.9 µm Reprosil C18 particle (Dr Maisch GmbH). The peptides were eluted with a gradient from 7 to 25% B phase in 75 min then to 90% B in 15 min, at flow rate of 300 nL/min. The mobile phase compositions were, water with 0.1 % formic acid (A) and 80% acetonitrile 0.1% formic acid (B). The temperature of the column was maintained at 60ºC using a column oven. The tandem mass spectra were acquired with higher-energy C-trap dissociation (HCD) of the 10 most intense ions (m/z 300–2000, charge states > 1+). The MS1 resolution was set to 120,000, with 3 x 10^6 AGC target value and a maximal injection time of 100 ms. MS/MS spectra were acquired in the Orbitrap with a resolution of 15,000 (at m/z 400), a target value of 50,000 ions, a maximum injection time of 250 ms. Dynamic exclusion was set to 30 s. Triplicate analysis were performed for all samples in randomized batches.

Peptide and protein identification and quantification

The mass spectrometry raw files were processed using MaxQuant software version 1.5.5.1 (Cox and Mann, 2008). Uniprot human database (May 2017) was used to search the peptide

data using Andromeda (Cox et al., 2011) as a search algorithm. The search parameters specified trypsin digestion with a maximum of two missed cleavages, carbamidomethylation of cysteine as fixed term modification and N-terminal acetylation and methionine oxidation as variable modifications. The peptide and protein level false discovery rates (FDR) were set to 0.01. The match between the runs option was enabled to transform the identifications across the mass spectrometric measurements. The label free quantification method (MaxLFQ) was used to determine the relative intensity values of proteins and to normalize the protein intensities between the samples (Cox et al., 2014). Prior to the downstream data-analysis, data was filtered to remove proteins with less than two unique peptides. Contaminants and reverse hits were also removed (Table S1). The proteomic mass spectrometry data presented in this paper were submitted to PRIDE (Vizcaíno et al., 2016) and have the accession number PXD008973.

Proteomics data analysis

All data analyses were performed using the R statistical programming software environment version 3.4.3 (R Core Team, 2015).

Exploratory data analysis

To explore the similarity of the samples and the grouping of the biological replicates in the LFQ-normalized data, the R-package pheatmap (Kolde, 2015) was used (Supplementary Figure 2). Pearson correlation coefficient was used as a similarity measure and hierarchical clustering with complete linkage for clustering the samples.

Differential expression analysis

The Reproducibility Optimized Test Statistic (ROTS) (Elo et al., 2008; Suomi et al., 2017) was used to detect the DE proteins between the conditions. Differential expression was examined separately for each comparison and time point. The examined comparisons were Th0 – Thp at 24h, Th0 – Thp at 72h, Th17 – Thp at 24h, Th17 – Thp at 72h, Th17 – Th0 at 24h and Th17 – Th0 at 72h. Technical replicates for a biological replicate were averaged and the data was log2-transformed prior to the differential expression analysis. FDR of 0.05 was used as a threshold to define the DE proteins. Differentially expressed proteins whose logarithmic fold change (LogFC) was > 0, were considered as up-regulated and proteins whose LogFC was < 0 were considered as down-regulated. Z-score standardization of the DE proteins in the compared samples was used for visualizing the changes in expression with heatmaps.

Enrichment analysis

The enriched GO biological processes were identified using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (Huang et al., 2009a, 2009b). The GO FAT terms, which filter out the broadest terms, were considered. The enrichment analysis was performed using the DE proteins over both time points as the input and the whole detected and filtered proteome as the background reference. A biological process was considered enriched if it had FDR ≤ 0.05.

The enrichment of molecular types and cellular locations was further examined using the Ingenuity Pathway Analysis (IPA) (QIAGEN Inc.) (Krämer et al., 2014). The enrichment of cellular locations and protein types between Th17 and Th0 was examined at 24h and 72h using the time point specific DE proteins as input for IPA and the whole detected and filtered proteome as a background reference. All the resulting IPA location and type information was collected.

Protein-protein interactions

To investigate protein-protein interactions, the DE proteins over both time points (24h and 72h) were entered into the STRING functional protein association networks database (Szklarczyk et al., 2017). Both predicted and known high confidence (interaction score ≥ 0.7) interactions were considered. The resulting interaction network was imported into Cytoscape version 3.6.0 (Shannon et al., 2003) for visualization. Clusters in the interaction data were identified using the Cytoscape plug-in clusterMaker v2 (Morris et al., 2011) with the granularity parameter (inflation value) set to 1.8 as suggested by (Brohée and van Helden, 2006). The GO biological process enrichment results over both time points were used to calculate the most frequent terms for each cluster, from which the most representative biological processes for each cluster were selected. The proteins in the main cluster were grouped according to the representative processes.

To further examine functionality in the identified network, enriched biological pathways in the main cluster (cluster 1) was examined using the Cytoscape plug-in ReactomeFIViz version 7.0.1 (Wu et al., 2014). The ReactomeFIViz-plugin was used to discover subclusters in cluster 1 and to identify the enriched biological pathways in the discovered subclusters. ReactomeFIVIz examines enriched pathways in a gene/protein set against multiple pathway-databases (Wu et al., 2014)]. An FDR of 0.01 was used as a threshold for biological pathways in the subclusters to discover only the most important functionalities.

Targeted Proteomics Validation

Selected reaction monitoring (SRM) mass spectrometry was used to validate the relative abundance of ATP1B1, PALLD, ACSL4, FHOD1, SMTN and RDX in the Th17 cells at 72 h. Heavy-labeled synthetic peptides (lysine $^{13}C_6$ $^{15}N_2$ and arginine $^{13}C_6$ $^{15}N_4$) peptides were obtained for the targets of interest (Thermo Fischer Scientific) and were selected on the

basis of their stability, consistency and intensity in the discovery data. For these validations, four additional cultures were prepared from the cord blood of four donors.

Skyline software (MacLean et al., 2010) was used to evaluate the top five most intense transitions from the MS/MS spectra of the heavy labelled standard peptides and assess the relative performance of the native peptides in the spiked validation samples.

The samples were prepared using the same FASP digestion and desalting protocols used for discovery. These were then spiked with synthetic heavy labelled analogues of the peptide targets and a retention time standard (MSRT1, Sigma) for scheduled selected reaction monitoring. The LC-MS/MS analyses were conducted using Easy-nLC 1000 liquid chromatograph (Thermo Scientific) coupled to a TSQ Vantage Triple Quadrupole Mass Spectrometer (Thermo Scientific). The column configuration included a 20 x 0.1 mm i.d. pre-column in conjunction with a 150 mm x 75 µm i.d. analytical column, both packed with 5 µm Reprosil C_{18}-bonded silica (Dr Maisch GmbH). A separation gradient from 8% to 43% B in 27 min, then to 100% B in 3 min, was used at a flow rate of 300 nl/min (the mobile phase compositions are as indicate above). The raw SRM data are available through PASSEL (Farrah et al., 2012) with the dataset identifier PASS01204

Skyline was used to select the transition used for the assays subsequently process the data generated. The MSStats (3.8.4) plugin included in the Skyline software was used for the group comparison between cases and controls. The summed intensities of GAPDH peptides were used as a global standard to normalize the data from each analysis and Tukey's median polish method was used as the summary method.

Transcriptomics data analysis

RNAseq sample preparation

RNA samples from five biological replicates derived from Th0 and Th17 cultures of five individual donors were collected at 72h time point. RNA was isolated (RNeasy Mini Kit,

QIAGEN) and DNase treated (RNase-Free Dnase Set; QIAGEN). Library preparation was performed according to Illumina TruSeq® Stranded mRNA Sample Preparation Guide (part # 15031047). RNA-seq with 50 nt read length was performed at the Finnish Functional Genomics Centre (FFGC) with HiSeq 3000 instrument using TruSeq chemistry and the raw data was base called with CASAVA1.8.

Preprocessing of the raw data

The RNA-seq raw reads were mapped to the Ensembl human reference genome GRCh38 (Genome Reference Consortium Human Build 38) (Zerbino et al., 2018) using the STAR aligner (Dobin et al., 2013) version 2.5.2. The read counts were generated using the featureCounts tool (Liao et al., 2014) in the Subread software package (Liao et al., 2013) version 1.5.1. Uniquely mapped reads were used for further analysis. The uniquely mapped reads were filtered for lowly expressed genes (genes with counts per million (cpm) >1 in at least 5 replicate samples were retained) and used for further analysis. The filtered gene counts were normalized using the trimmed mean of M-values (TMM) normalization from the Bioconductor package edgeR (McCarthy et al., 2012; Robinson et al., 2010) after which the data was transformed to counts per million (cpm), offsetted by 1 and log2-transformed. Differential expression analysis was performed similarly to proteomics using ROTS (Elo et al., 2008; Suomi et al., 2017). An FDR of 0.05 was used as a threshold to define the DE genes. The RNAseq data presented in this study was submitted to the Gene Expression Omnibus (GEO) and has the Series record GSE118974.

Comparison between human and mouse proteomics data

To compare differentially regulated proteins (differentially expressed proteins and proteins detected in only one condition) between Th17 and Th0 in human and mouse at 72h during Th17 polarization, we used the published mouse proteomics raw data (Mohammad et al., 2018) and pre-processed it similarly to the human data using MaxQuant (version 1.5.5.1)

with LFQ-normalization, filtering out proteins with less than two unique peptides, and removing contaminants and reverse hits. Similarly, as with the human proteomics data, the differential expression analysis was performed using ROTS. To make the comparison of the differentially regulated proteins more comprehensive, we used a threshold of FDR 0.1 in both datasets to define the DE proteins. Mouse genes related to proteins were mapped to orthologous human genes for the comparison using Ensembl BioMart (Zerbino et al., 2018). All the orthologous mouse genes from Ensembl 92 database to the human reference genome GRCh38 were considered. If multiple orthologous human genes existed for a given mouse gene, the most similar human orthologous gene according to the Ensembl database was selected.

Immunoblot Analysis

For the western blot analysis, cell samples were lysed in either RIPA (Pierce, #89901) or Triton-X buffer (50 mM Tris-HCl, pH 7.5; 150 mM NaCl; 0.5% Triton-X-100; 5% glycerol; 1% SDS), supplemented with proteinase (Roche) and phosphate inhibitors (Roche) on ice with vortex every 5 min. Lysed samples were sonicated for 5 min under ice cold conditions (Bioruptor UCD-200; Diagenode), followed by centrifugation at maximum speed for 20 min at 4°C. Supernatants containing proteins were collected and quantified (DC Protein Assay; Bio-Rad). The samples containing 15–30 μg of protein was boiled with 6x sample loading dye (330 mM Tris-HCl, pH 6.8; 330 mM SDS; 6% β-ME; 170 mM bromophenol blue; 30% glycerol). Samples were loaded on 4–20% gradient SDS-PAGE gels (Biorad), transferred to nitrocellulose membranes (Bio-Rad), and probed with the antibodies listed in Supplemental Table 7. During the course of this study, immuno-blotting membranes were either cut into sections and each section was probed with separate antibodies, or the blots were striped with striping buffer (25mM Glycine and 1%SDS; pH 2.5) and re-probed successively with different antibodies recognizing proteins with different molecular mass.

IL-17A secretion

IL-17A production is a characteristic of CD4+ T cells differentiated toward the Th17 phenotype and can be detected in cell-culture supernatant at 72h using either the Milliplex MAP human IL-17A kit (Merck Millipore; HCYTOMAG-60K-01), Bioplex Human IL-17A Cytokine/Chemokine 96-Well Plate Assay (Bio Rad; Cat. no. 171B5014M, 171304090M) or Human IL-17A Duoset ELISA kit (R&D Biosystems DY317-05, DY008). The amount of IL-17A secreted by Th17 cells was normalized with the number of living cells determined based on forward and side scattering in flow cytometric analysis (LSRII flow cytometer; BD Biosciences).

IFN-γ secretion

Culture supernatants from CD4+ T cells differentiated toward the Th17 phenotype at 72h were harvested and assayed by ELISA for IFN-γ secretion (Milliplex MAP human IL-17A+IFN-γ kit; Cat.no. HCYTOMAG-60K-02), according to the manufacturer's protocols. For appropriate analysis, all values below the detectable range were considered zero.

Flow cytometry

Flow cytometry analysis of cell-surface receptor CCR6 detection was done at 72h after initiation of Th17 cell priming. In short, cells were washed twice with PBS, and staining was done in staining buffer (0.5% FBS/0.1% Na-azide/PBS) for 15 min at 4°C and data is either acquired in flow cytometer on same day in staining buffer or cells were fixed with 1% formalin. LSRII flow cytometer (BD Biosciences) was used in data acquisition. Living cells were gated for analysis based on forward and side scattering. Detection of IFN-γ producing cells was determined by intracellular cytokine staining with anti-IFN-γ-FITC (BD Biosciences). Briefly, cells were stimulated with phorbol 12-myristate 13-acetate and ionomycin (PMA) for 5 h. GolgiStop (BD Biosciences) was added after 2 h, and cells were fixed in 4% paraformaldehyde solution. Fixed cells were stained with fluorescent antibodies

in 0.1% saponin permeabilization buffer and analysed on a LSRII (BD Biosciences). For OASL and ATF3 staining, cells were first fixed with 4% paraformaldehyde and permeabilized with Perm III buffer (BD Biosciences) before analysis. Cells were incubated with primary antibodies, followed by 30 min incubations with the secondary antibody when needed. The staining was controlled by using isotype specific antibodies or only with the secondary antibodies. The information regarding the antibodies used for staining is listed in Supplemental Table 7.

Quantitative Real-time PCR

Total RNA was isolated using RNeasy kit (Qiagen, Valencia, CA). cDNA was synthesized with Reverse Transcription kit (Applied Biosystems, Foster City, CA) using oligo dT primers according to the manufacturer's instruction. TaqMan primers and probes for IL-17A, IL-17F, RORC and SATB1 were designed with Universal ProbeLibrary Assay Design Centre (Roche), in Absolute QPCR ROX Mix (Thermo Scientific). EF1a gene was used as endogenous control. The qPCR runs were analysed using the 7900HT Fast Real-Time PCR System (Applied Biosystems).

Cell transfection with small interfering RNA (siRNA)

Freshly isolated CD4+ T cells from umbilical cord blood were suspended in Optimem I (Invitrogen) and transfected with SATB1 targeting siRNA oligonucleotide (Sigma) (Table S7) using the nucleofection technique (Lonza). Four million cells were transfected with 5ug of siRNA after which the cells were rested at 37^0 C for 24h in RPMI 1640 medium (Sigma-Aldrich) supplemented with pen/strep, 2 mM L-glutamine and 10% FCS ($2x10^6$ cells/ml) and subsequently activated and cultured under Th17 culturing condition as described above.

Statistical analysis

A two-tailed student’s t-test was used for determining the statistical significance of IL17A and IFN-y secretion and percentage of CCR6 and IFN-y expressing cells at 72h of culture

from three to five independent cultures. Statistical analysis of the mass spectrometry data is described in the respective method.

Supplemental References

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J. V, and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Cox, J., Hein, M.Y., Luber, C. a, and Paron, I. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. … *13*, 2513–25

Vizcaíno, J.A., Csordas, A., Del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. *44*, D447–D456.

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. Vienna Austria.

Kolde, R. (2015). pheatmap : Pretty Heatmaps. R Packag. Version 1.0.8.

Elo, L.L., Filén, S., Lahesmaa, R., and Aittokallio, T. (2008). Reproducibility-optimized test statistic for ranking genes in microarray studies. In IEEE/ACM Transactions on Computational Biology and Bioinformatics, p.

Suomi, T., Seyednasrollah, F., Jaakkola, M.K., Faux, T., and Elo, L.L. (2017). ROTS: An R package for reproducibility-optimized statistical testing. PLoS Comput. Biol. *13*.

Huang, D.W., Sherman, B.T., Lempicki, R.A., Huang, D., Sherman, B., Lempicki, R., Huang, D., Sherman, B., Lempicki, R., Maere, S., et al. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc.

Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. Bioinformatics.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res.

Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., and Ferrin, T.E. (2011). ClusterMaker: A multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics.

Brohée, S., and van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics.

Wu, G., Dawson, E., Duong, A., Haw, R., and Stein, L. (2014). ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. F1000Research.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: An open source document editor for creating and analyzing targeted proteomics experiments.

Bioinformatics

Farrah, T., Deutsch, E.W., Kreisberg, R., Sun, Z., Campbell, D.S., Mendoza, L., Kusebauch, U., Brusniak, M.Y., Hüttenhain, R., Schiess, R., et al. (2012). PASSEL: The PeptideAtlas SRMexperiment library. Proteomics.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res.

Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics.

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

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

Mohammad, I., Nousiainen, K., Bhosale, S.D., Starskaia, I., Moulder, R., Rokka, A., Cheng, F., Mohanasundaram, P., Eriksson, J.E., Goodlett, D.R., et al. (2018). Quantitative proteomic characterization and comparison of T helper 17 and induced regulatory T cells. PLOS Biol.