2. Methods

2.1 Cytokine induced polarisation of human immune cells

2.1.1 Isolation of human $CD4^+$ T cells and monocytes from peripheral blood

Blood from healthy donors was obtained as leukocyte reduction cones from the NHS Blood and Transplant, Cambridge, UK. Upon collection, 10 ml of blood was diluted in 50 ml of RPMI-1640 culture medium supplemented with 10% foetal calf serum (FCS, Sigma Aldrich) and 1 mM EDTA (Sigma Aldrich) to avoid clotting. Each 25 ml of blood in RPMI was layered on top of 17 ml of Ficoll-Paque (GE Life Sciences) and gradient centrifugation was performed at 830 g for 20 minutes without break. Peripheral blood mononuclear cells (PBMCs) were recovered from the middle layer using a pasteur pipette and washed twice in 50 ml FACS buffer (Dulbecco's Phosphate Buffer Saline (PBS, Sigma), supplemented with 1mM EDTA and 2% FCS). The cells were then resuspended at 50 x 10^6 cells/ml.

Approximately 100 x 10^6 PBMCs were transferred to a 15 ml tube (Thermo Scientific) for monocyte isolation. This procedure was carried out via positive magnetic beads cell selection using the Human $CD14^+$ Positive Selection Kit (EasySep™, STEMCELL Technologies) following the manufacturer's instructions. The remaining PBMCs were used for $CD4^+$ T cell enrichment, which was carried out via negative magnetic beads cell selection using the Human $CD4^+$ T Cell Enrichment Kit (EasySep™, STEMCELL Technologies). Samples enriched in $CD4^+$ cells were resuspended at 100 x 10^6 cells/ml and incubated for 30 minutes with FACS antibodies against the following surface markers: CD4, CD25, CD127, and CD45RA (**Table 2.1**). Finally, cells were washed in 2 ml FACS buffer. For each fluorophore, single stain compensation controls were prepared using 1x10^6 cells from the same sample for each individual stain. Shortly before sorting, cells were incubated with a 1:3000 dilution of the nucleic acid stain 4',6-diamidino-2-phenylindole (DAPI) (Sigma Aldrich) for 5 minutes in order to label dead cells.

Table 2.1. Product description of the FACS antibodies used throughout this thesis

Antibody	Fluorophore	Manufacturer	Clone	Volume per 100 µl sample
Anti-human CD4	APC	BioLegend	OKT4	3 µl
Anti-human CD25	PE	BioLegend	M-A251	8 µl
Anti-human CD127	FITC	eBioscience, Thermo Scientific	RDR5	3 µl
Anti-human CD45RA	BV785	BioLegend	HI100	6 µl
Anti-human CD69	PE-Cyanine7	eBioscience, Thermo Scientific	FN50	4 µl
Anti-T-bet MAb	PE-Cyanine7	eBioscience, Thermo Scientific	4B10	5 µl
Anti-Gata-3 MAb	PE-Cyanine7	eBioscience, Thermo Scientific	TWAJ	5 µl
Anti-FoxP3 MAb	BV421	BioLegend	206D	5 µl
Anti-human CD14	Alexa Fluor 647	BD Pharmigen	MphilP9	2 µl
Anti-human CD68	PE-Cyanine7	BioLegend	Y1/82A	2 µl
Anti- human MerTK	BV421	BioLegend	590H11G1E3	5 µl
Anti-human HLA-DR	PE	BioLegend	L243	5 µl

Fluorescence-activated cell sorting (FACS) was performed on the samples in a MoFlo XDP sorter (Beckman Coulter). Naïve T cells, defined as $DAPI^{-}$ $CD4^{+}$ $CD25^{low}$ $CD127^{high}$ $CD45RA^{+}$, were recovered. Cell sorting was performed by the Cytometry Facility of the Wellcome Trust Sanger Institute. Following sorting, naïve T cells were washed in 50 ml FACS buffer and resuspended at $2x10^{6}$ cells/ml in RPMI-1640 medium supplemented with 10% FCS, 2 mM L-glutamine (Sigma-Aldrich), and 1X Penicillin-Streptomycin (Gibco™, Thermo Scientific) (further referred to as complete RPMI).

2.1.2 Activation and cytokine polarisation of $CD4^+$ naïve T cells

Human T-activator CD3/CD28 dynabeads (Gibco™, Thermo Scientific) were washed and diluted in complete RPMI medium. Naïve T cells were plated in a 24-well plate (Corning) at $1x10^6$ cells per well in 2 ml volume, and supplemented with dynabead solution when appropriate. Unless otherwise stated, the ratio of cells to dynabeads was kept at 2:1. The control samples were supplemented with complete RPMI medium. Where required, polarising cytokines were added in combinations and concentrations shown in **Table 2.2**. All cytokines used in this study were purchased from PeproTech, while all antibodies used for polarisations were obtained from R&D. Cells were incubated at 37˚C for 16 hours or 5 days, respectively.

Activation beads were removed after either 16 hours or five days of stimulation. In order to do so, the cell suspension for each condition was transferred into a 1.5 ml Eppendorf tube. Tubes were then incubated for five minutes in a DynaMag™ magnet (Thermo Scientific) for 1.5 ml tubes, after which the media containing cells was carefully transferred to new Eppendorf tubes without touching the wall closest to the magnet in order not to transfer beads.

After beads removal, 20 µl of cells were mixed with 20 µl of Cellometer ViaStain™ AOPI (Nexcelom Bioscience) solution to assess their viability and counted in an automated cell counter (Cellometer Auto 2000™ Cell Viability Counter). Additionally, for a selection of samples, cells were also counted in a flow cytometer using AccuCheck™ counting beads.

2.1.4 Differentiation and stimulation of monocyte derived macrophages

Isolated $CD14^+$ cells were seeded into 6-well culture plates (Corning), 2.5×10^5 cells in 2 ml complete RPMI per well. For macrophage differentiation, the media was supplemented with 50 ng/ml M-CSF (PeproTech). Cultures were incubated for 7 days at 37 ˚C.

Following 7 days of culture, the media was supplemented with cytokine (PeproTech) polarisation cocktails (**Table 2.3)**. After 6 hours, the supernatant containing the cytokines was removed and cells were washed twice with PBS.

Table 2.2 Composition of $CD4^+$ naïve T cells cytokine-polarisation cocktails

POLARISING CONDITION	ACTIVATION BEADS RATIO	CYTOKINE COMBINATION	CYTOKINE CONCENTRATIONS
Unstimulated	--	No cytokines	---
Th0	1:2	No cytokines	--
Th1	1:2	IL-12	50 ng/ml
		Anti-human IL-4 antibody	1 ng/ml
Th2	1:2	IL-4	10 ng/ml
		Anti-human IFN-γ antibody	1 ng/ml
Th17	1:2	IL-6	50 ng/ml
		IL-23	20 ng/ml
		IL-1β	10 ng/ml
		TGF-β1	5 ng/ml
		Anti-human IL-4 antibody	1 ng/ml
		Anti-human IFN-γ antibody	1 ng/ml
iTreg	1:2	TGF-β1	5 ng/ml
TNF-α	1:2	TNF-α	20 ng/ml
IL-10	1:2	IL-10	50 ng/ml
IL-21	1:2	IL-21	50 ng/ml
IFN-β	1:2	IFN-β	10 ng/ml
IL-21	1:2	IL-21	20 ng/ml

Table 2.3. Composition of macrophage cytokine-polarisation cocktails

POLARISING CONDITION	CYTOKINE COCKTAIL	CONCENTRATION
Unstimulated (M0)	No cytokines	---
M1	IFN-γ	50 ng/ml
M2	IL-4	50 ng/ml
TNF-α	TNF-α	50 ng/ml
IL-23	IL-23	20 ng/ml
IL-26	IL-26	20 ng/ml

2.1.5 Flow cytometry

All the samples described below were analysed with a LSR Fortessa (BD) FACS analysers.

2.1.5.1 Assessing T cell activation

Following 16 hours and five days of T cell stimulation, cells were transferred into a 96-well plate (Corning), washed once with 200 µl FACS buffer, and incubated with FACS antibodies against CD69 and CD25 for 20 minutes. Cells were then incubated with 1:3000 DAPI solution for 5 minutes and washed twice with FACS buffer to remove unbound antibody and DAPI. Compensation controls for each fluorophore were prepared by incubating UltraComp eBeads™ compensation beads (Thermo Scientific) with 1 µl of the respective antibody.

2.1.5.2 Immunophenotyping of monocyte derived macrophages

Following seven days of differentiation, monocyte derived macrophages were detached from the plate by incubating them with 1X Non-enzymatic Cell Dissociation Solution (Sigma Aldrich). Cells were resuspended in 1 X Fix/Perm Buffer (BioLegend) containing Human TruStain FcX (BioLegend) Fc receptor blocking solution for 10 minutes. After FcR blocking, the cells were incubated for 30 minutes at room temperature with antibodies against CD14, CD68, MerTK and HLA-DR. (**Table 2.1**). Cells were then washed in 1 ml 1 X Fix/Perm Buffer and resuspended in 200 µl FACS buffer.

2.1.5.3 Immunophenotyping of polarised $CD4^+$ T cells

After five days of cytokine polarising stimulation, T cells from the Th1, Th2, Th17 and iTreg conditions were phenotyped for expression of relevant transcription factors. Cells were fixed and permeabilized using the FOXP3 Fix/Perm Buffer Set (BioLegend), according to the manufacturer's instructions. Permeabilized cells were incubated for 30 minutes with FACS antibodies against the transcription factors T-bet, GATA3, and FoxP3. Each of these stainings was performed separately and independently across all conditions.

2.1.5.4 Data analysis

All data obtained by flow cytometry was analysed with FlowJo™ version 10 (TreeStar). FlowJo was used to quantify the relative percentages and median fluorescence intensities (MFI) of cell populations. The results from this analysis were then imported into the R programming environment (version 3.1.1), where they were visualized as scatter plots and boxplots using ggplot2 library.

2.2 Analysis of gene expression in cytokine polarised cells

2.2.1 Isolation of RNA from T cells and monocyte derived macrophages

Following polarisation of monocyte derived macrophages and naive $CD4^+$ T cells, total RNA was isolated from all conditions by resuspending 3 x 10^5 cells in 500 µl of TRIzol™ (Invitrogen, Thermo Scientific). Samples were stored at -80˚C until RNA extraction was performed. Thawed cell suspension was transferred to a 1.5 ml MaXtract High Density tube (Qiagen), and 100 µl of chloroform (Sigma Aldrich) was added to each sample. The tubes were vigorously mixed for 15 seconds and incubated at room temperature for 10 minutes. Following incubation, samples were centrifuged at 21,000 g for 5 minutes at 4˚C. The 300 µl of aqueous phase from each sample was used for RNA isolation with the RNeasy™ MinElute Cleanup Kit (Qiagen) as instructed in the manufacturer's handbook.

2.2.2 Library preparation and RNA-sequencing

The quality of total RNA was assessed using RNA 6000 Nano Chip (Agilent Technologies) on a 2100 Bioanalyzer Instrument (Agilent Technologies) according to the manufacturer's instructions. Only samples with an RNA integrity number (RIN) higher than eight were processed for library preparation, which was performed by the Sequencing Pipelines Facility at The Sanger Institute, following the TruSeq mRNA Sample Preparation Guide by Illumina

(136). Briefly, 1 µg total RNA was used for mRNA purification using poly-T oligo attached magnetic beads in two rounds of washing and elution. The resulting mRNA was transferred to a PCR tube and fragmented under high temperature. The fragmented mRNA was resuspended in RNAse free water, and the first cDNA strand was synthesised using random primers and reverse transcriptase. The second cDNA strand was then synthesised using RNAse to remove the RNA template. The overhangs were converted into blunt ends using and “End Repair Mix” and the blunt ends were adenylated using the Klenow fragment and dATP in an A-tailing buffer. Following A-tailing, index adapters were ligated to the ends of the DNA, and the cDNA was amplified by PCR, generating a library suitable for sequencing. The adapters used contained barcode sequences which allowed multiplexing of the 56 samples in the same flow cell. The libraries were finally spread across two lanes of an Illumina HiSeq 2500 for clustering and sequencing.

2.2.3 RNA-seq data analysis

Read mapping

Sequencing data was aligned, demultiplexed and quality assessed by the Sequencing Facility at The Sanger Institute. It was released in CRAM format. Once receiving the data, the CRAM files were converted to FASTQ using BioBamBam release 2.0.8 (137). Next, FASTQ files with the same barcodes were merged. Reads were mapped to the hg38 release of the Human Reference Genome using the Spliced-Transcripts Alignment to a Reference (STAR) software version 2.5.3 (138) with the default parameters. The output of STAR mapping, obtained in a Binary Alignment Map (BAM) format, was filtered for high-quality reads using SAMtools version 1.3.1 (139). Only reads with a minimum mapping quality (MAPQ) of 20 were kept for further analysis. Gene expression quantification was estimated using featureCounts version 1.22.2 (140).

Quality control and exploratory data analysis

To reduce sources of variability unrelated to the study, counts from non protein-coding genes, the HLA region on chromosome 6 (chr6:25,000,000-47,825,000), and the Y chromosome were removed. Next, genes showing low expression levels, defined as a sum of counts lower than 20 (considering all samples), were also removed from the analysis. A table with sample information was built, which included: cell type, polarising condition assayed, time point, and batch number. Principal component analysis (PCA) (141, 142) was performed using DESeq2 (135). To stabilize the variance of lowly expressed genes, the count data was transformed using the regularized logarithmic transformation before

performing PCA (135). Throughout this thesis, DESeq2 version 1.16.1 (Bioconductor release 3.5) was used.

Finally, the Euclidean distance between samples was calculated and used to perform hierarchical clustering and build heatmaps using “pheatmap” R package (143). Batch correction was performed using a linear model with the batch correction function included in the “limma” R package (144). R version 3.1.1 is used throughout this study.

Statistical power estimation

In order to estimate the statistical power of the study across several comparisons, we divided the data into three smaller data sets: macrophages (12 samples), CD4+ T cells 16 hours (22 samples) and CD4+ T cells 5 days (22 samples). These same data sets were later use for analysing differential gene expression. DESeq2 was used to estimate the coefficient of variation (CV) of each gene in the three data sets. To do so, the CV of each gene was defined as approximately equal to the squared root of its dispersion, as suggested by the package developers (135). Next, the global CV of each group was estimated as the median CV of all genes.

Following CV estimation, we modelled the statistical power as a function of gene coverage. Coverage was defined as the number of raw counts detected for each gene (134). For this calculation, we used the R package “RNASeqPower”, which estimates power using a generalised linear model (GLM) based on the Negative Binomial (NB) distribution (134). Throughout these estimations, we considered a sample size of 2 and assumed an average effect size (fold change) of 2, and an FDR of 0.05.

Finally, we used “RNASeqPower” (134) to estimate the global power of the study. We did this separately for each of the three sample groups. First, we obtained the median gene coverage using the counts() function of DESeq2 (135) and used this median value to calculate the power of detecting genes with a fold change of 2 at an FDR of 0.05, given the CV and sample size of the study.

Differential expression analysis

The RNA counts table was used to assess differential gene expression in R. Separate analyses were performed for T cells and macrophages across different time points. To analyse transcriptional changes due to cytokine polarisation of monocyte derived macrophages, RNA counts were compared against unstimulated cells. Conversely, to analyse transcriptional changes upon cytokine polarisation in $CD4^+$ T cells, Th0 cells were used as a baseline for the comparison in order to identify only differences explained by cytokine polarisation as opposed to T cell activation. For this, P values were computed using Wald's Test and the negative binomial (NB) distribution in DESeq2, with gene size and sequencing depth correction (135, 145). The Benjamini-Hochberg (BH) method was applied for multiple testing correction (146). All genes with a BH-adjusted P value ≤ 0.05 and an absolute log2 fold change ≥ 1 were classified as differentially expressed. When loading the data into DESeq2, the following linear models were used to represent the null and alternative hypotheses:

$$H_0: \quad \mathbf{Y} = \mathbf{B_0} + \mathbf{B_1 D} + \boldsymbol{\varepsilon}$$

$$H_A: \quad \mathbf{Y} = \mathbf{B_0} + \mathbf{B_1 D} + \mathbf{B_2 C} + \boldsymbol{\varepsilon}$$

Where **Y** is a vector containing the observed RNA counts, $\mathbf{B_0}$ an intercept term (mean of counts), $\mathbf{B_1}$ and $\mathbf{B_2}$ the linear model coefficients, **D** the categorical variable batch number, **C** the cytokine-polarising condition (treatment), and **ε** a random error term.

Functional annotation of differentially expressed genes and overrepresentation analysis (OA)

Up and down regulated genes were functionally annotated using gene ontology (GO) (147, 148). Upregulation was defined as a log2 fold change ≥ 1, and downregulation as a log2 fold change ≤ -1. The R package gProfileR (149) was used to annotate each gene to its respective cellular component (CC), molecular function (MF), and biological process (BP) GO terms. Next, an overrepresentation analysis of GO terms was performed using gProfileR. The observed proportion of genes assigned to each term was compared to the expected proportion, assuming a hypergeometric distribution (149). The complete list of genes used for differential expression analysis (14,399 protein-coding genes) was used as a background, and the hypergeometric P values were corrected for multiple testing using Benjamini-Hochberg's FDR method (146).

Gene co-expression network analysis

To analyse shared gene regulation in response to cytokine stimulation, correlation networks were inferred. To do this, a variation from the standard workflow for weighted gene co-expression network analysis (150) was implemented. The Pearson correlation coefficient of each pair of genes across all samples was selected as a measure of similarity between genes and calculated as specified below. Firstly, a correlation matrix was built using gene expression data from all the samples of interest. The correlation matrix was defined as follows:

$$\mathbf{X} = \begin{pmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \\ \vdots \\ \mathrm{X}_n \end{pmatrix} = \begin{pmatrix} & \text{Samples} & & & \\ \text{Genes} & x_{1\,1} & x_{1\,2} & \cdots & x_{1\,m} \\ & x_{2\,1} & x_{2\,2} & \cdots & x_{2\,m} \\ & \vdots & \vdots & \ddots & \vdots \\ & x_{n\,1} & x_{n\,2} & \cdots & x_{n\,m} \end{pmatrix}$$

$$\boldsymbol{\rho} = \begin{pmatrix} & \text{Genes} & & & \\ \text{Genes} & 1 & \rho_{1\,2} & \cdots & \rho_{1\,n} \\ & \rho_{2\,1} & 1 & \cdots & \rho_{2\,n} \\ & \vdots & \vdots & \ddots & \vdots \\ & \rho_{n\,1} & \rho_{n\,2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \frac{cov(\mathrm{X}_1, \mathrm{X}_1)}{\sigma_{\mathrm{X}_1} * \sigma_{\mathrm{X}_1}} & \frac{cov(\mathrm{X}_1, \mathrm{X}_2)}{\sigma_{\mathrm{X}_1} * \sigma_{\mathrm{X}_2}} & \cdots & \frac{cov(\mathrm{X}_1, \mathrm{X}_n)}{\sigma_{\mathrm{X}_1} * \sigma_{\mathrm{X}_n}} \\ \frac{cov(\mathrm{X}_2, \mathrm{X}_1)}{\sigma_{\mathrm{X}_2} * \sigma_{\mathrm{X}_1}} & \frac{cov(\mathrm{X}_2, \mathrm{X}_2)}{\sigma_{\mathrm{X}_2} * \sigma_{\mathrm{X}_2}} & \cdots & \frac{cov(\mathrm{X}_2, \mathrm{X}_n)}{\sigma_{\mathrm{X}_2} * \sigma_{\mathrm{X}_n}} \\ \vdots & \vdots & & \vdots \\ \frac{cov(\mathrm{X}_n, \mathrm{X}_1)}{\sigma_{\mathrm{X}_n} * \sigma_{\mathrm{X}_1}} & \frac{cov(\mathrm{X}_n, \mathrm{X}_2)}{\sigma_{\mathrm{X}_n} * \sigma_{\mathrm{X}_2}} & \ddots & \frac{cov(\mathrm{X}_n, \mathrm{X}_n)}{\sigma_{\mathrm{X}_n} * \sigma_{\mathrm{X}_n}} \\ & & \cdots & \end{pmatrix}$$

With **X** being the expression matrix containing regularized log-transformed counts of the 14,399 protein-coding genes across all samples, and **ρ** a matrix containing the Pearson correlation coefficients of gene pairs. We then proceeded to find the main co-expression partners of the genes of interest. To define the co-expression network, the gene of interest was set to be the central node and a signum adjacency function with threshold (150, 151) was used to determine its interactors. This adjacency function was defined as follows:

$$\boldsymbol{a}_{i,j}(\boldsymbol{\rho}_{i,j}) = \begin{cases} 1 & if\ \rho_{i,j} \geq \tau \\ 0 & if\ \rho_{i,j} < \tau \end{cases}$$

With $\mathbf{a_{ij}}$ equaling 1 when the two genes **i** and **j** belong to the same co-expression network, and 0 when the two genes do not show a sufficiently strong co-expression pattern. For consistency of the network topology, the network size rather than value of **τ** was kept constant, as suggested in the literature (150, 151). To achieve this, the correlation coefficients of the gene of interest with all the remaining genes were extracted from the correlation matrix and approximated by a normal distribution. The threshold was defined as the correlation coefficient equivalent to the 99th percentile of this distribution. In sum, genes with the 1% highest correlation coefficients with the gene of interest were defined as its interactors.

Genes within each co-expression network of interest were then tested for common transcriptional regulation. To do this, TRANSFAC (149, 152) was used to perform an overrepresentation analysis for transcription factor binding sites (TFBS) with the same statistical models and corrections described above. Shared TFBS were ranked by FDR-adjusted P value.

2.3 Protein expression analysis of activated T cells

2.3.1 Isolation and quantification of protein from T cells

For proteome quantification, unstimulated T cells, Th0 and Th1 cells were used in two technical replicates. Only three cell states were selected in order to test the technique's reproducibility and optimise the experimental protocols, with the plan on profiling all studied conditions in the future.

Pellets from $3x10^6$ cells were washed twice with PBS, the supernatant removed, and dried cell pellets were stored at -20˚C until protein isolation was performed. For protein quantification, the cell pellets were thawed and resuspended in 150 µl 0.1 M triethylammonium bicarbonate (TEAB) buffer (Sigma Aldrich) supplemented with 0.1% SDS. Pulse probe sonication was performed with 40% power (EpiShear™) on ice for 20 seconds, after which the samples were boiled for 10 minutes at 96˚C in a heat block. This procedure was performed twice. Lysed cells were then centrifuged at 12,000 rpm for 10 minutes in order to remove cellular debris, and protein quantification was performed using the Quick Start Bradford Protein Assay (Bio-Rad) as specified by the manufacturer's instructions. Protein samples were then divided into 100 µg aliquots.

2.3.2 Protein digestion, TMT labeling, and HPLC

Protein digestion, TMT labelling, and HPLC analysis were performed by Marta Baldrighi in collaboration with Theodoros Roumeliotis and the proteomics facility at The Sanger Institute, as specified in the initial collaboration note to this dissertation. Briefly, protein aliquots were resuspended in a 5 mM solution of tris-2-carboxymethyl phosphine (TCEP) buffer (Sigma Aldrich) and incubated for 1 hour at 60˚C in order to reduce the disulfide bonds. Iodoacetamide (IAA) was added to the sample and adjusted to a final concentration of 10 mM, in which the sample was incubated for 30 minutes at room temperature, in the dark. MS grade Pierce Trypsin (Thermo Scientific) was added at a mass ratio of 1:30, and the samples were incubated overnight for peptide digestion.

The digested protein samples were diluted to a total volume of 100 µl in 0.1 M TEAB buffer. A volume of 41 µl anhydrous acetonitrile was added to each vial of TMT reagents (Thermo Scientific) and vortexed. The content of each TMT vial was then transferred to the corresponding protein sample. After 1 hour, the reaction was quenched by adding 8 µl of 5% hydroxylamine. All the protein samples were combined into a single tube and dried using a speedvac concentrator. Samples were stored at -20˚C until fractionation could be performed.

High pH Reverse Phase (RP) peptide fractionation was performed with the Waters XBridge C18 column (2.1 x 150 mm, 3.5 µm, 120 Ä) on a Dionex™ UltiMate 3000 HPLC system (Thermo Scientific). A 0.1% solution of ammonium hydroxide was used as mobile phase A, while mobile phase B was composed of 99.9% acetonitrile with 0.1% ammonium hydroxide. The TMT-labelled samples were reconstituted in 100 µl mobile phase A, centrifuged and injected into the column, which operated at 0.2 ml/min. The fractions collected from the column were dried with the SpeedVac concentrator and stored at -20˚C until MS could be performed.

2.3.3 LC-MS/MS

Liquid Chromatography-Mass Spectrometry (LC-MS) was performed using a Dionex™ UltiMate 3000 HPLC system (Thermo Scientific) coupled with the Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific). This was performed by the Proteomics Facility at The Sanger Institute.

Dried samples were reconstituted in 40 µl 0.1% formic acid, of which 7 µl were loaded to the Acclaim PepMap 100 trapping column (100 µm x 2 cm, C18, 5 µl, 100 Ä) at a flow rate of 10 µl/min. Then, multi-step gradient elution was performed at 45˚C using the Dionex™ Acclaim PepMap RSLC capillary column (75 µm x 50 cm, 2 µm, 100 Ä). A 0.1% solution of formic acid was used as mobile phase A, and a 80% acetonitrile, 0.1% formic acid solution as mobile phase B. Precursors with mass resolution of 120k, AGC 3 x 105 and IT 100 ms were isolated for CID fragmentation with quadrupole isolation width of 1.2 Th. Collision energy was set at 35%. Furthermore, MS3 quantification spectra were acquired via further fragmentation for the top 10 most abundant CID fragments.

2.3.4 Proteomics data analysis

Protein identification and quantification

Protein discovery analysis was performed by Theodoros Roumeliotis and the proteomics facility at The Sanger Institute. Briefly, the obtained mass spectra were submitted to Sequest HT search in Proteome Discoverer™ 2.1 (Thermo Scientific). Spectra were searched for tryptic peptides with maximum 2 miscleavage events and a minimum length of 6 amino acids. The TMT 6-plex at N-terminus were defined as static modifications. Peptide confidence was estimated with the Percolator node, and the FDR was set at 0.01 and validation was based on q-values. Spectra were then searched against 20,165 reviewed human entries of UniProt. TMT quantification was performed with the Reporter Ion Quantifier node, and a window tolerance of 15 ppm. Only peptides belonging to protein groups were used for quantification.

Quality control

Raw protein counts were normalised to the sample median. Next, the data was imported to the R programing environment and a $\log_2$ transformation was applied to improve data visualization. PCA was performed on the transformed protein counts using the “prcomp” package of base R.

RNA-protein correlation analysis

Raw RNA counts and median normalised protein counts were imported to R, and a $\log_2$ transformation was applied to both data sets. Next, the RNA-seq biological replicates and MS technical replicates were averaged, and the Pearson correlation between both molecular traits was calculated. Finally, the $\log_2$ fold change between unstimulated T cells and each polarising condition (Th0 and Th1) were calculated for both data sets, and a Pearson correlation coefficient was computed.