4. Analysis of gene expression in cytokine polarised cell states

4.1 Overview

Throughout this study we sought to delineate the differences between cytokine induced cell states by studying their gene expression. In this chapter, I describe the results of a low coverage RNA-seq study of $CD4^+$ T cells polarised to four known lineages (Th1, Th2, Th17 and iTreg) and the additional five conditions treated with different cytokines linked to autoimmunity (IL-10, IL-21, IL-27, IFN-β, and TNF-α). These cells were assessed after 16 hours and 5 days of stimulation. Additionally, I describe the results from RNA-seq performed on macrophages polarised for six hours towards the classical M1 and M2 phenotypes, as well as with the additional conditions with three different cytokines linked to autoimmunity (IL-23, IL-26, and TNFα). Finally, I describe the results of differential gene expression analysis followed by pathway enrichment and gene co-expression network analysis.

4. 2 Quality control of sequencing data

Quality assessment was performed by the Sequencing Facility at The Sanger Institute (**Table 4.1** and **4.2**). We observed a median fragment size of 177 bp. Approximately 91% fragments matched with the human genome. The GC fraction was close to the expected value. Furthermore, we observed that multiplexing barcodes were efficiently decoded, with an average 98% perfect matches, and only 0.83% sequences with no match. We also observed high accuracy throughout the sequencing run, as determined by quality scores throughout cycles (**Figure 4.1**). We concluded that sequencing had been performed appropriately, generating high quality data.

Table 4.1 Global quality metrics of two lanes sequenced run in an Illumina HiSeq 2500

Lane Number	Number of Cycles	Tag Decode Rate (%)	Tag Decode CV	Median Insert Size	Top Reference Matches (%)
1	158	99.17	7.06	177	*Homo sapiens* 91.6 *Pan troglodytes* 84.6
2	158	99.16	7.10	177	*Homo sapiens* 91.9 *Pan troglodytes* 84.6

Table 4.2 Read specific quality metrics of two lanes sequenced in an Illumina HiSeq 2500

Lane Number	Read	Adapters (%)	GC Fraction (%)	Yield (kb above Q20)	Average mismatch (%)
1	Forward	0.06	48.0	11,123,259	2.56
	Reverse	0.03	49.1	10,965,772	2.52
2	Forward	0.06	48.0	11,160,621	2.52
	Reverse	0.03	49.1	11,013,167	2.46

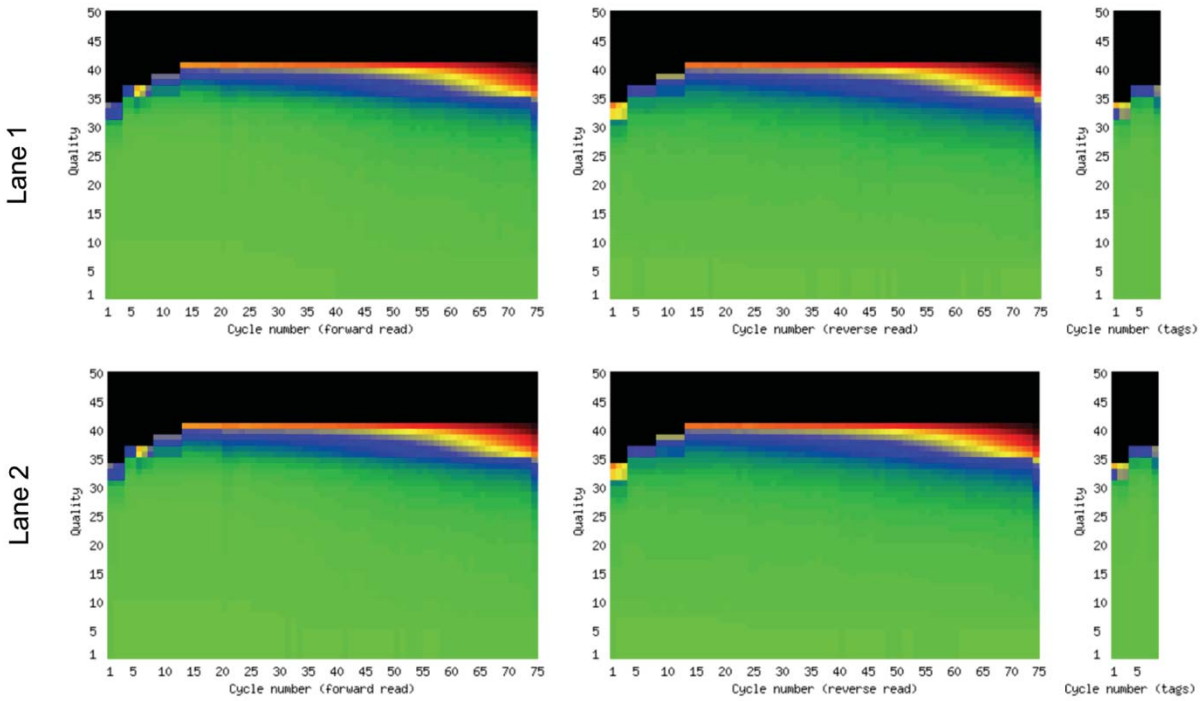

Figure 4.1 Sequencing quality is high throughout the run Following library preparation, the mRNA library was spread across two lanes of an Illumina HiSeq 2500 for clustering and sequencing. The average quality score at each cycle was plotted for the forward and reverse reads, as well as the tags. Each sequencing lane was analysed separately.

Following quality control, we re-mapped the 5.25 x 10^6 paired-end reads per sample using STAR (138). Approximately 89.5% of reads were uniquely mapped. Next, we used featureCounts (140) and obtained 83% of reads assigned to features. We concluded that mapping efficiency was high.

To estimate the reproducibility, we compared the expression values of two replicates of the same condition. This analysis was done using biological replicates, since the study included no technical replicates. We transformed the data using the log2 function. We observed that for three selected conditions the log2 counts were highly correlated upon visual inspection (**Figure 4.2**). Hence, we calculated the Pearson correlation for all pairs of replicates and observed an average correlation coefficient higher than 0.96 (**Table 4.3**). We concluded that the RNA-seq results were reproducible.

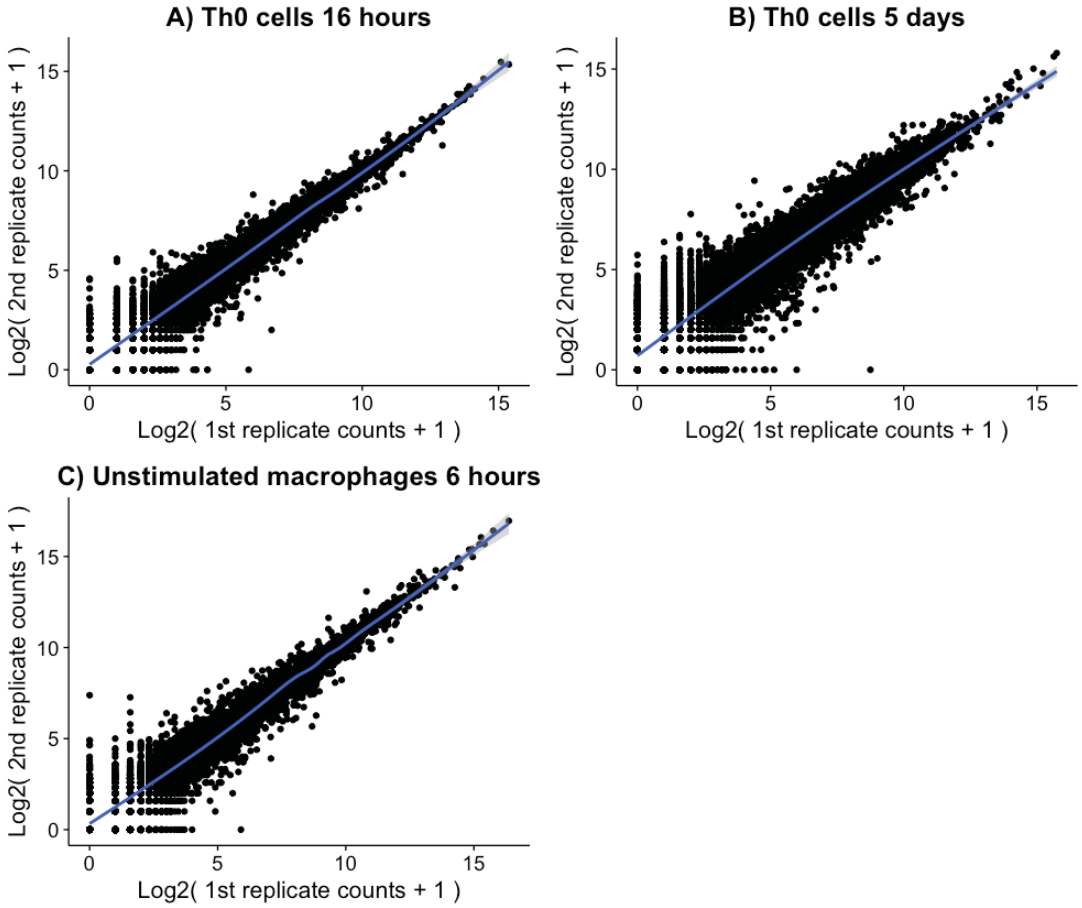

Figure 4.2 RNA-seq results are reproducible Following mapping and summarisation, the RNA counts of two biological replicates of three representative were plotted against each other. The samples were selected so as to include all cell types and time points and log2 transformed counts were used.

Table 4.3 Correlation coefficient of log transformed RNA counts for the two biological replicates of each condition.

Cell Type	Time Point	Condition	Pearson Correlation
T cell	16 hours	Unstimulated	0.977
		Th0	0.981
		Th1	0.981
		Th2	0.981
		Th17	0.981
		iTreg	0.981
		IL-10	0.981
		IL-21	0.981
		IL-27	0.981
		IFN-β	0.982
		TNF-α	0.981
	5 days	Unstimulated	0.963
		Th0	0.958
		Th1	0.963
		Th2	0.966
		Th17	0.961
		iTreg	0.966
		IL-10	0.967
		IL-21	0.962
		IL-27	0.965
		IFN- β	0.971
		TNF-α	0.965
Macrophage	6 hours	Unstimulated	0.978
		M1	0.98
		M2	0.979
		IL-23	0.971
		IL-26	0.977
		TNF-α	0.977

4. 3 Estimation of statistical power

To determine how powered the study was, we estimated the statistical power to detect genes with 2-fold change differential expression at an FDR of 0.05. We divided the samples in three groups according to cell type and time point and used DESeq2 and “RNASeqPower” (134) to model power as a function of the expression level. We observed that power increased for genes with higher counts (**Figure 4.3**) to a similar extent in all groups. On average, power was larger than 0.6 for genes with raw counts larger than 20, and above 0.8 for genes with raw counts higher than 50. We concluded that the study might be powered for detecting differences in genes with expression levels of 50 counts or more.

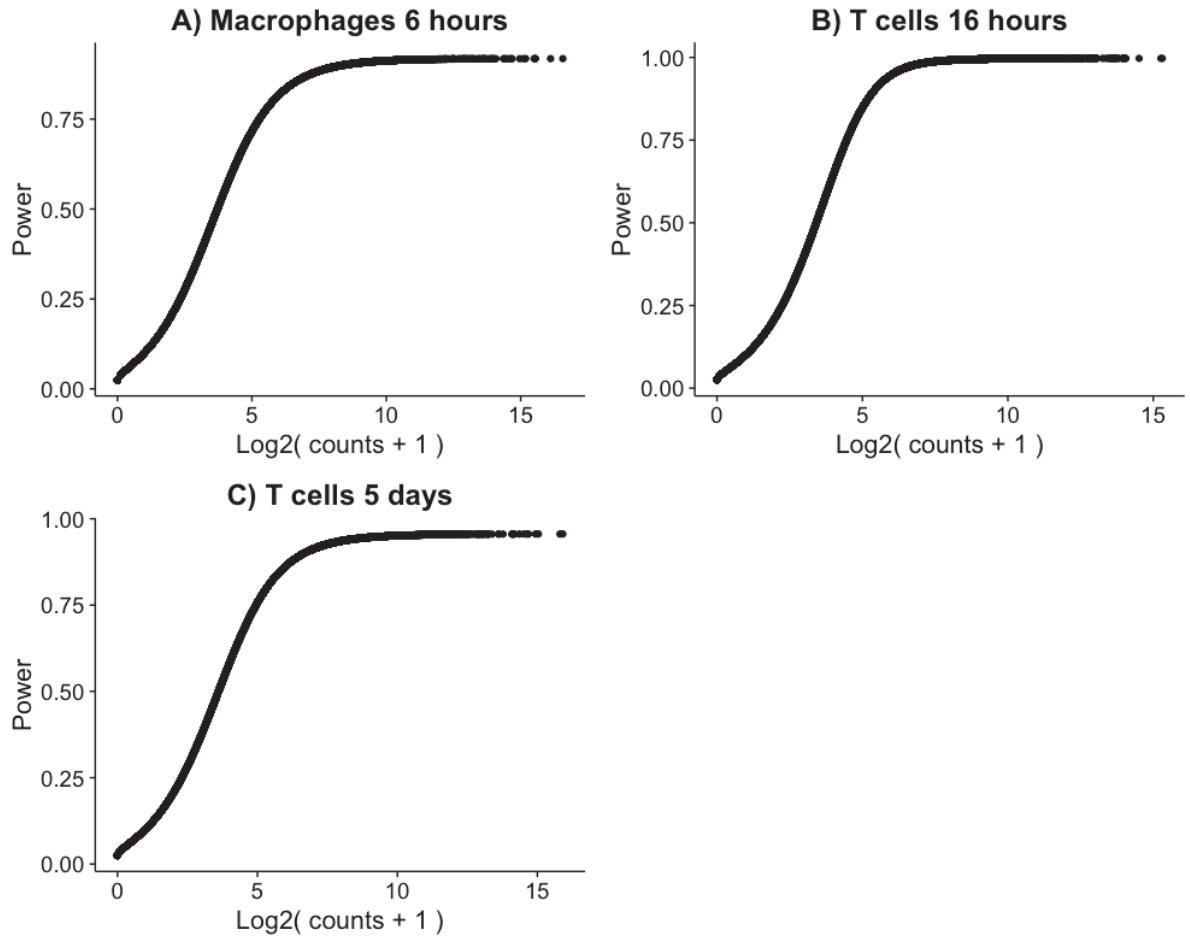

Figure 4.3 Statistical power increases for highly expressed genes Gene expression data was divided in three groups, and DESeq2 was used to estimate their coefficient of variation. This coefficient was used to model the statistical power for detecting a gene as a function of its coverage (RNA count) using “RNASeqPower”. For this calculation, we specified an FDR of 0.05, a sample size of 2, and a fold change (effect size) of 2. The power to detect each gene was plotted against its log2-transformed counts.

New we estimated the average statistical power per group. We used DESeq2 to compute the overall median of RNA counts for each group and concluded that, on average, the study had 0.8 power to detect genes with a fold change of 2 at 0.05 FDR (**Table 4.4**).

Table 4.3 Average power for detecting genes with a fold change of 2 in each group of samples.

Cell type	Time point	Median of Counts	Median CV	Fold Change	Average Power
T cell	16 hours	53	0.147	2	0.93
	5 days	73	0.189	2	0.88
Macrophage	6 hours	59	0.207	2	0.81

4.4 Exploratory data analysis

To explore the data in more detail, we used DESeq2 (135) to perform data transformation and PCA in all samples of the data set. We observed a clear clustering by cell type, with PC1 separating monocytes from T cells (**Figure 4.4A**). This separation accounted for 83% of the variance. On the other hand, PC2 explained 12% of the variance and separated the two different time points within the $CD4^+$ T cell samples. We recapitulated these results by performing PCA on T cell samples only (**Figure 4.4C**). Here, PC1 accounted for 75% of the variance and separated cells stimulated for 16 hours from cells stimulated for five days. We also observed a separate cluster containing four samples, these conditions corresponded to unstimulated cells at both time points. Therefore, PC2 might separate resting cells from the remaining conditions.

To obtain more detailed insights into the data, we repeated the PCA separately for each cell type. We observed that PC1 separated M1 macrophages from the rest of the samples (**Figure 4.4B**) and explained more than 50% of the variance. This analysis also showed separation of the unstimulated, M2, TNF and IL26 macrophage conditions, with both biological replicates clustering tightly together. However, macrophage samples stimulated with IL-23 did not cluster together in the PCA plot.

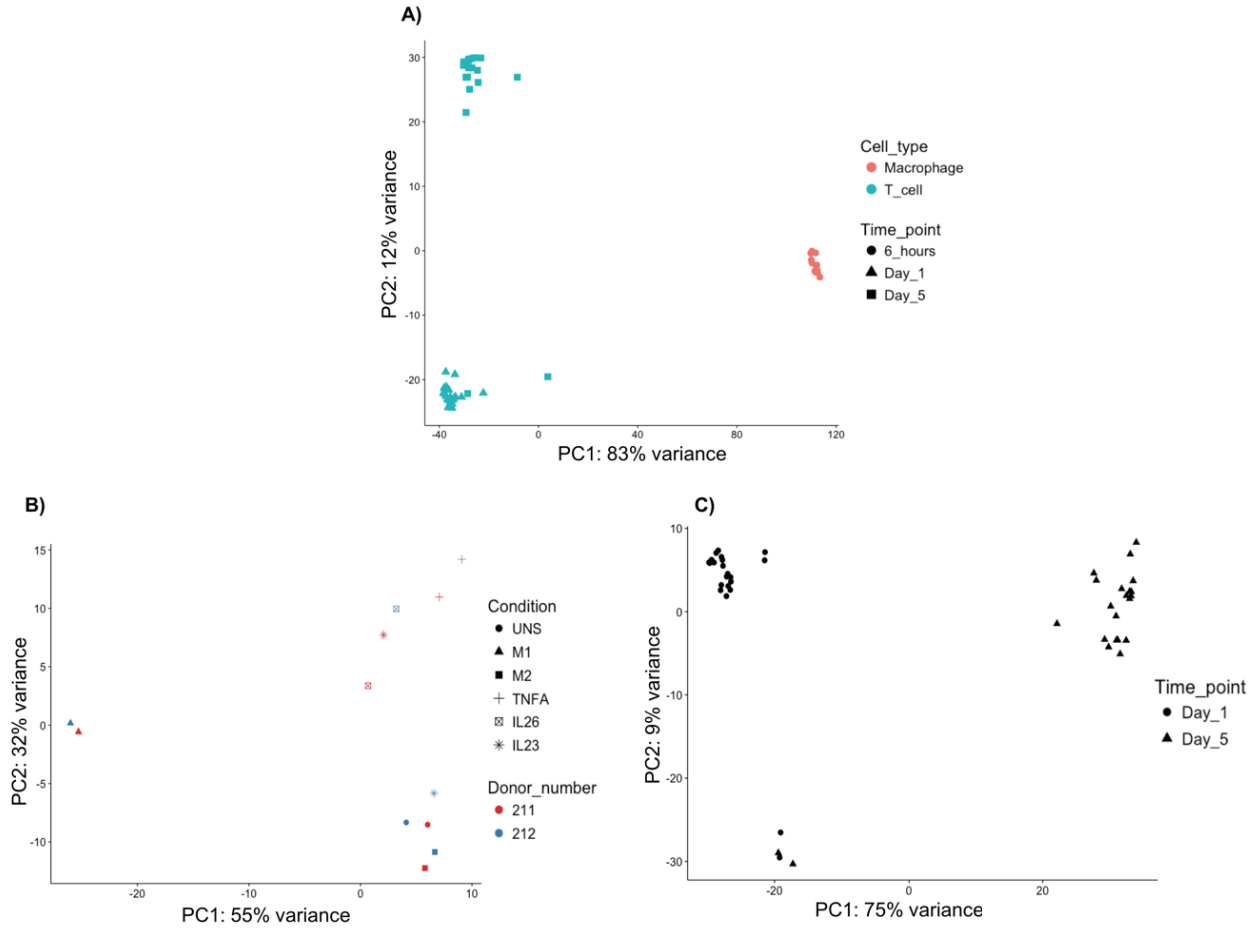

Figure 4.4 PCA separates different cell types, stimulation time points and conditions. PCA was performed on RNA counts from: **A)** All samples in the study, with colours represent different cell types and shapes different stimulation times. **B)** Monocyte derived macrophages from two independent biological replicates, with different shapes represent cytokine polarising conditions and colours biological replicates. **C)** $CD4^+$ T cells from two independent biological replicates, with shapes representing stimulation times. Here, the four unstimulated T cell samples clustered separately to the rest of the conditions near the bottom of the plot.

Next, we assessed the differences between polarisations in $CD4^+$ T cells. We removed the unstimulated samples from the analysis and asked whether samples separated by condition. Using transformed RNA counts, we calculated the distance between samples and performed hierarchical clustering using DESeq2. We observed a batch effect between the two individuals, which were also processed on different days (**Figure 4.5A**). Thus, we removed the batch effect using limma (144) and repeated the distance calculation and hierarchical clustering. We observed clustering of biological replicates in the dendrogram at both time points (**Figure 4.5B**). This suggested the presence of different transcriptional responses to different cytokine polarisations. However, we found that several conditions clustered with the control samples. Specifically, iTreg, TNFα, IL-10 and IL-23 were close to the Th0 control at

16 hours. The same was true for IL-10, TNF-α, IL-27 and IL-21 at five days. On the other hand, IFN-β, Th2, and IL-27, clustered separately to the rest of the samples at 16 hours. This was also true for Th1, Th2, Th17 and iTreg cells at five days.

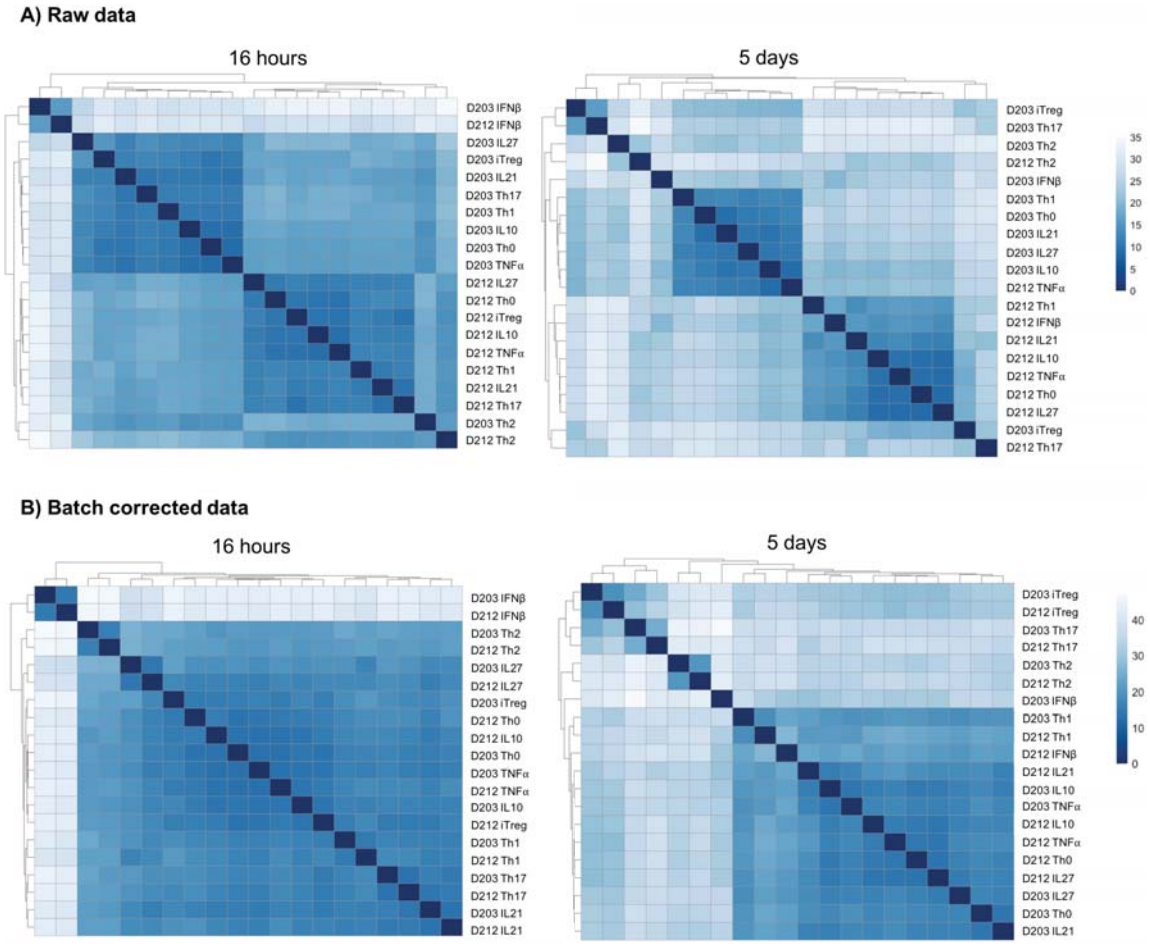

Figure 4.5 Batch correction improves the separation of polarising conditions in CD4$^+$ T cells Calculation of Euclidean distance between samples followed by hierarchical clustering was performed. Results are displayed as a heat map, with a colour gradient proportional to the distance between two samples. **A)** Samples from two different batches are shown. **B)** The same samples are shown after implementation of the batch correction. The labels indicate the batch number and the polarising condition of each sample.

We then asked whether transcriptional profiles agreed with descriptions of macrophage and T cell lineages in the literature. We analysed the RNA counts of the M1 macrophage markers IL-12, CXCL11, and COX-2 (94) and observed upregulation in the M1 samples as compared with M2 macrophages (**Figure 4.6A**). CXCL11 expression increased 7-fold in M1 macrophages. We repeated this analysis for the M2 markers COX-1 and MRC1 (94) and confirmed that their expression seemed higher in M2 than in M1 (**Figure 4.6B**). Notably, COX-1 and COX-2 seemed to have opposite expression patterns at the RNA level. Consequently, we concluded that polarisation with IFN-γ and IL4, respectively, induced the M1 and M2 polarisation in macrophages.

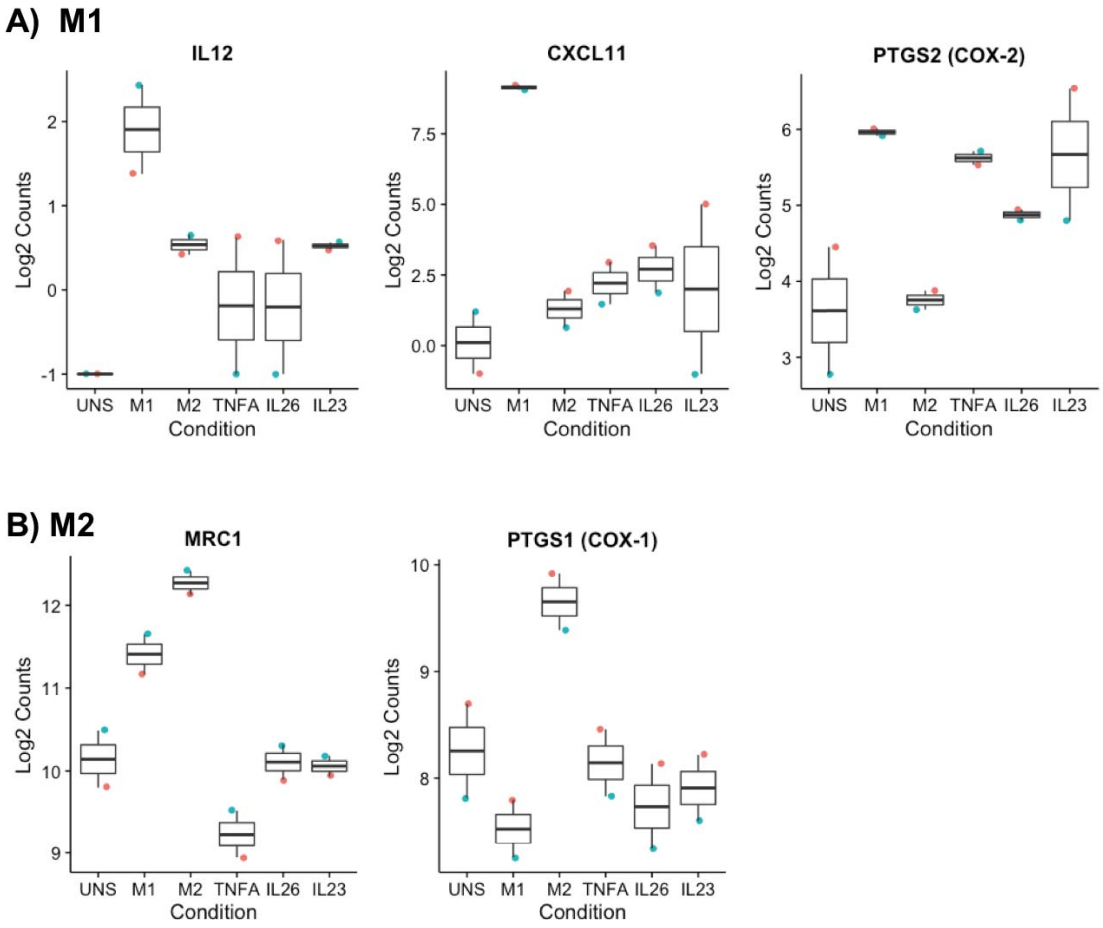

Figure 4.6 Lineage specific genes are expressed in M1 and M2 macrophages Bar plots of regularised log2 (rlog) values of RNA counts of macrophage specific. Different colours indicate two independent biological replicates. **A)** M1 macrophages were inspected for expression of IL12, CXCL11 and COX2. **B)** M2 macrophages were inspected for expression of MRC1 and COX1.

Next, we analysed the expression of lineage markers in $CD4^+$ T cells at five days. Th1 cells upregulated T-bet and IFN-γ relative to all other conditions (**Figure 4.7.1A**). In addition, we confirmed that Th2 cells (**Figure 4.7.1B**) expressed three fold more GATA-3 than other activated cells. We also observed expression of IL-4 and IL-13. Conversely, Th17 cells significantly downregulated GATA-3 but showed higher expression of RORC, IL17F, and CCR6 (**Figure 4.7.2C**). CCR6 was also upregulated by iTreg cells, which expressed two fold more FoxP3 than other lineages (**Figure 4.7.2D**). Moreover, we found upregulation of CTLA-4 in iTreg cells. The expression of the regulatory cytokine IL-10 did not change upon iTreg polarisation. We concluded that stimulation with cytokines for the Th1 (IL-12), Th2 (IL-4), Th17 (IL-23, TGF-β, IL-6), and iTreg (TGF-β) lineages effectively induced polarisation in $CD4^+$ cells.

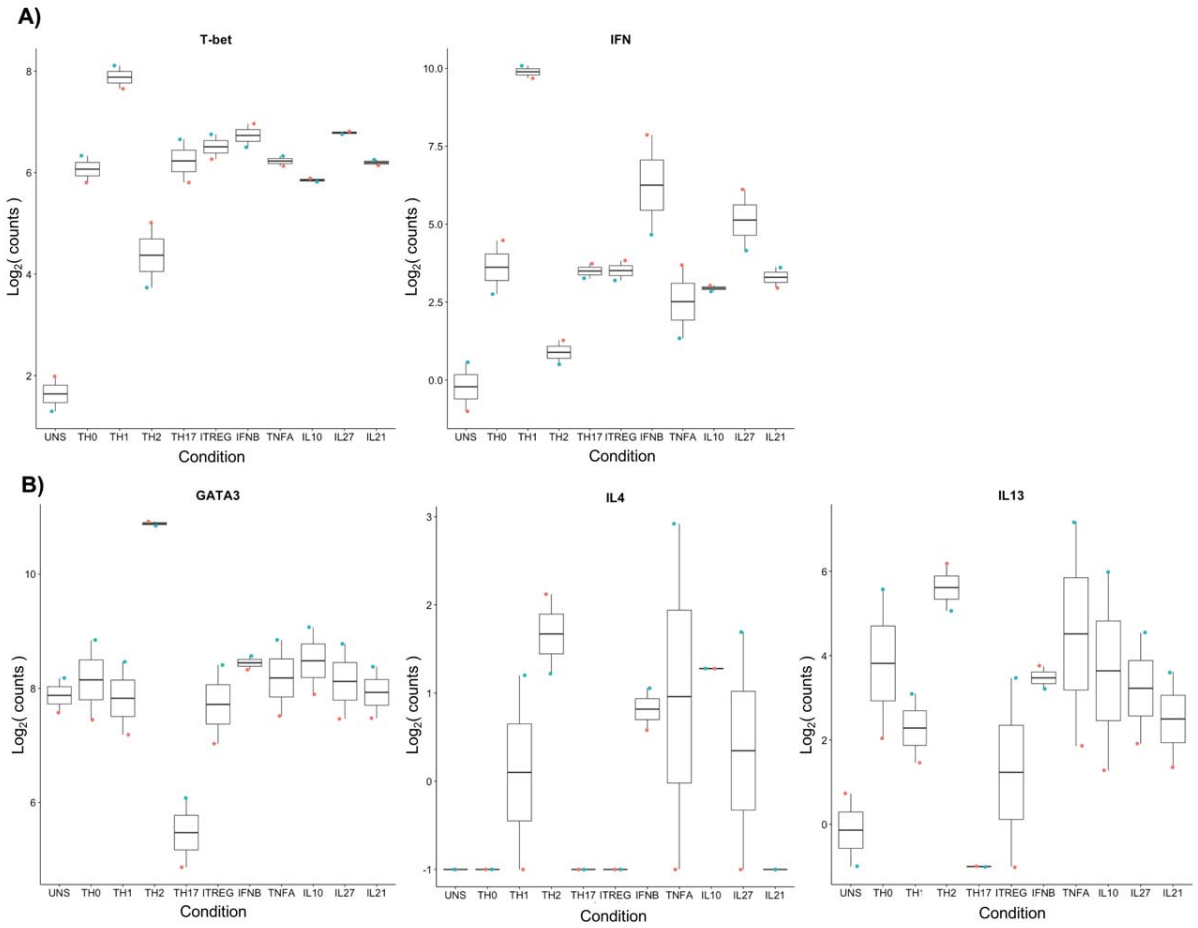

Figure 4.7.1 Hallmark genes are upregulated after T cell polarisation Bar plots of logarithmic values of RNA counts of T cell lineage specific genes. Different colours correspond to two independent biological replicates. Expression of **A)** T-bet and IFN-γ in Th1 and **B)** GATA3 and IL-4 in Th2.

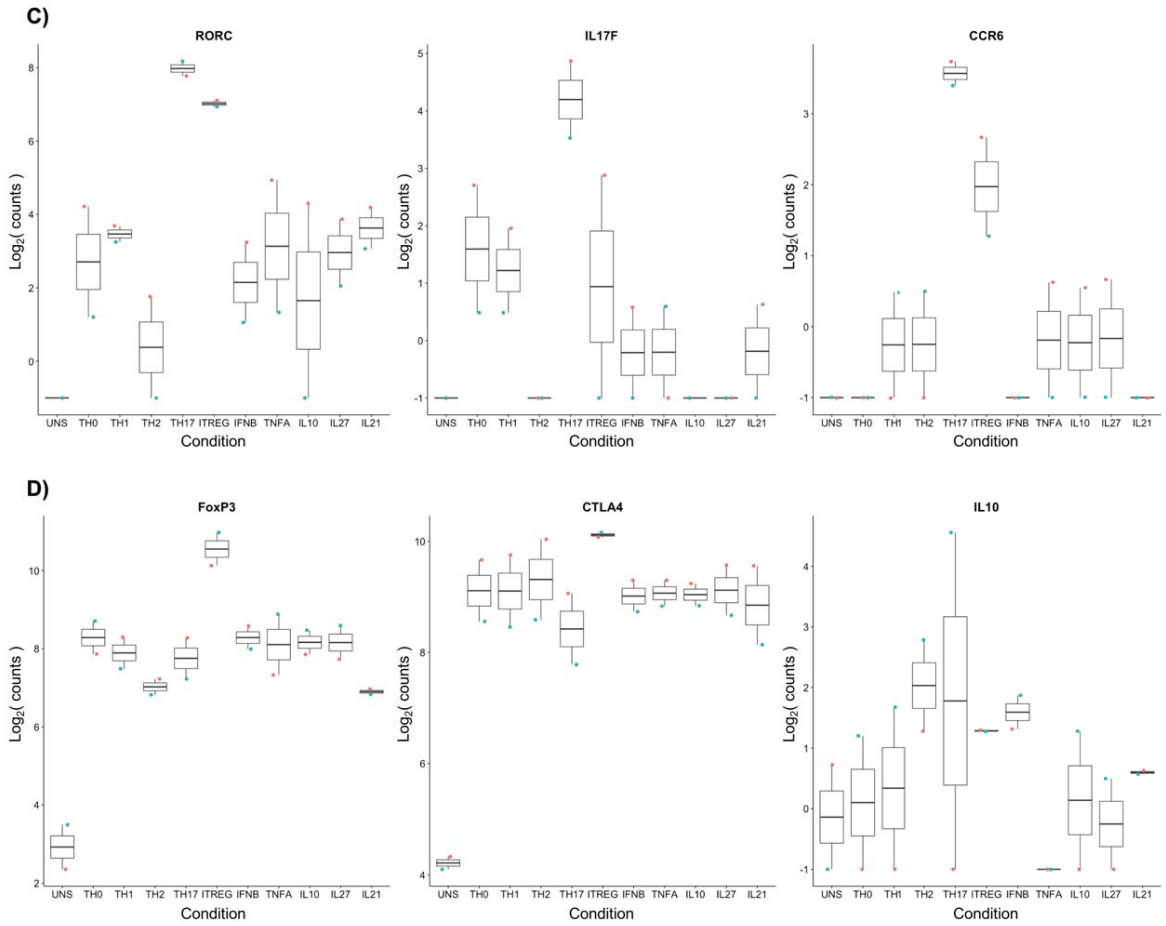

Figure 4.7.2 Hallmark genes are upregulated after T cell polarisation Bar plots of regularised log2 (rlog) values of RNA counts of T cell lineage specific genes. Different colours correspond to two independent biological replicates. Expression of **C)** IL-17F, RORC and CCR6 in Th17, and **D)** FoxP3 and CTLA-4 in iTreg cells.

4.5 Differential gene expression in polarised CD4$^+$ T cells

To identify genes specifically expressed in different CD4$^+$ T cell polarisation states we performed differential expression analysis. We did pairwise comparisons of RNA counts between each condition and non-polarised Th0 cells, which were used as negative control. Genes with an absolute $\log_2$ fold change larger than 1 and an adjusted P-value lower than 0.05 were considered significantly differentially expressed. At the 16-hour time point, we identified 140 and 29 differentially expressed genes in IFN-β and Th2 conditions, respectively. However, in the remaining conditions practically no genes reached statistical significance at this time point (**Figure 4.8A**). In contrast, after five days of polarisation we identified differential gene expression for all the lineages: 45 genes were differentially expressed in Th1, 361 in Th2, 489 in Th17, and 260 in iTreg cells (**Figure 4.8A**). However, we only found 63 differentially expressed genes upon IFN-β polarisation for five days.

Detailed information concerning differential expression analysis of these conditions, such as fold changes, P-values and mean gene counts, is presented at the end of this thesis (**Appendix**).

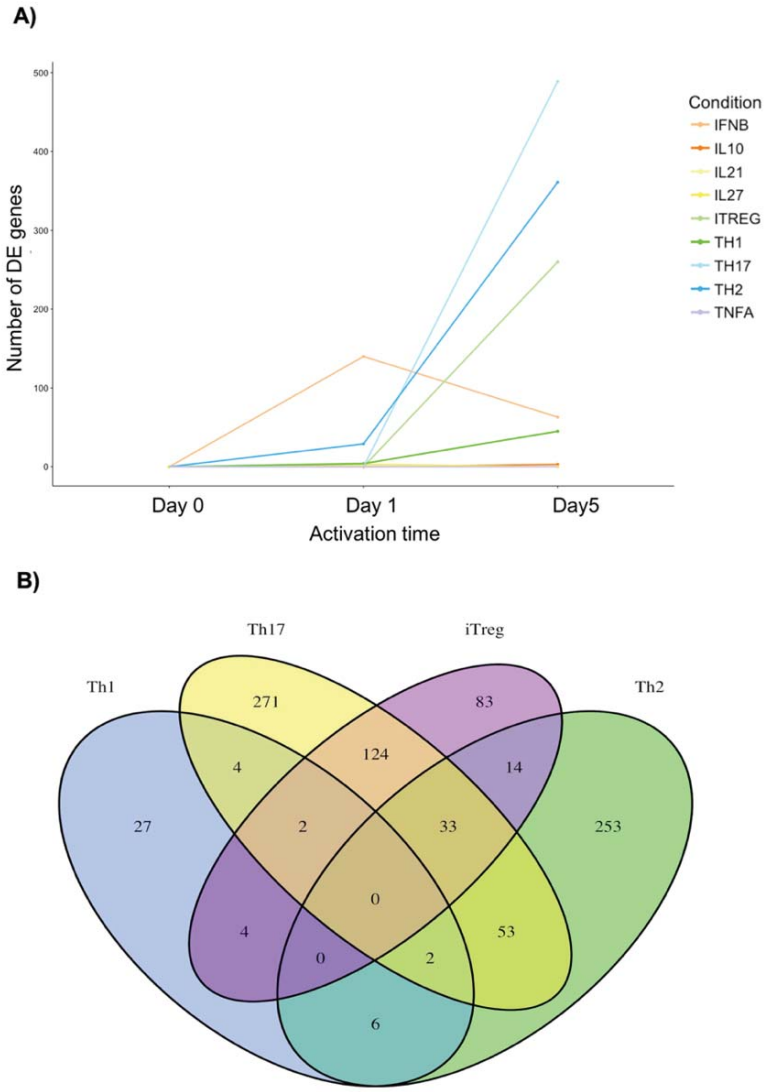

Figure 4.8 Transcriptional responses to cytokine induced polarisation are condition specific Differentially expressed genes were defined by comparing CD4+ T cells polarised with cytokines to unpolarised Th0 cells. Genes with an adjusted P-value ≤ 0.05 and an absolute $\log_2$ fold change ≥ 1 were considered differentially expressed. **A)** The numbers of differentially expressed genes per condition were plotted with respect to polarisation time, with each colour representing a condition. **B)** Overlap of differentially expressed genes across Th1, Th2, Th17 and iTreg cells at five days.

Next, we looked for shared differential gene expression across conditions. We overlapped the lists of differentially expressed genes obtained for all T cell conditions at day five and quantified the overlap. IFN-β stimulation was not included in this analysis since most changes appeared at a different time point and displayed entirely different transcriptional dynamics. No genes seemed to be shared between the conditions (**Figure 4.8B**). Th2 and Th17 appeared to have the most different transcriptional profiles, with 253 genes only present in Th2 and 271 only in Th17. On the contrary, iTreg shared approximately 60% of its transcriptional profile with Th17 cells, with only 83 apparently unshared genes. Furthermore, there were 27 genes with differential expression only detected in Th1 cells.

We hypothesised that responses to different cytokines have different transcriptional dynamics. To test this, we analysed the time trajectory of each differentially expressed gene in IFN-β and Th2 polarisation, the condition with the most differentially expressed genes (**Figure 4.9**). We observed substantial, stable expression changes between the first and fifth day in Th2 polarisation, with two large blocks of genes showing up and downregulation, respectively. On the contrary, most genes of the IFN-β response seemed to increase at 16 hours but to decrease at five days. To test for pathway enrichment, we analysed differential expression in Th1, Th2, Th17 and iTreg polarisations at the 5-day time point, while using the 16-hour time point for IFN-β stimulation.

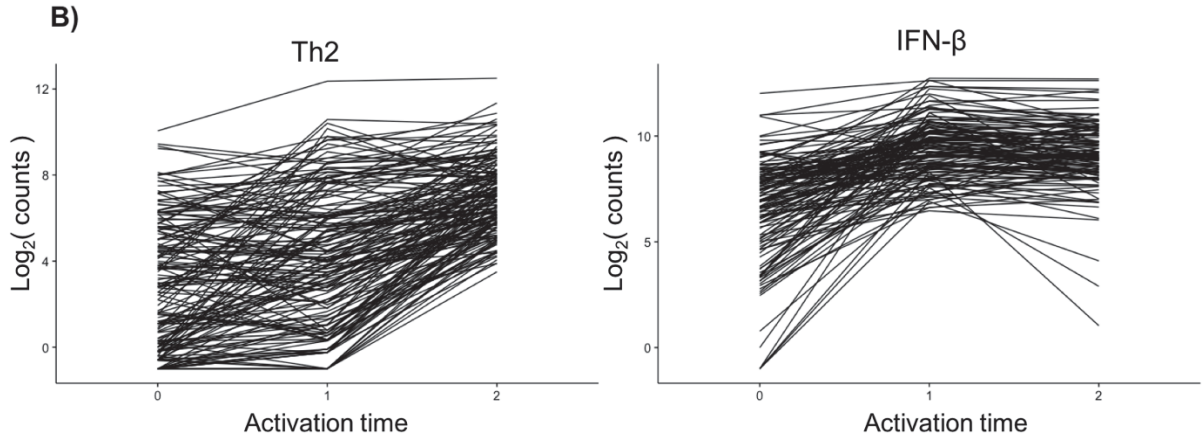

Figure 4.9 Different transcriptional dynamics between different induced T cell statesB) The time trajectory of each differentially expressed gene was plotted for Th2 cells and IFNβ stimulated cells. Regularised log2 (rlog) transformation was used.

Next, we functionally annotated the differentially expressed genes in each condition. First, we classified genes as upregulated or downregulated based on their $\log_2$ fold change. Using gProfileR (149), we annotated the genes in each group with GO terms for the cellular component (CC), molecular function (MF), and biological process (BP) categories (147). Next, we analysed which GO terms were overrepresented. We confirmed that the genes upregulated after 16 hours of IFN-β stimulation were enriched in interferon signalling pathways (**Figure 4.10**). Furthermore, their molecular function was mostly related to RNA binding. The genes driving this enrichment included interferon induced protein kinases and helicases, and 2'-5'-oligoadenylate synthetases (OAS). All of these terms characterise processes involved in the antiviral response (167, 168).

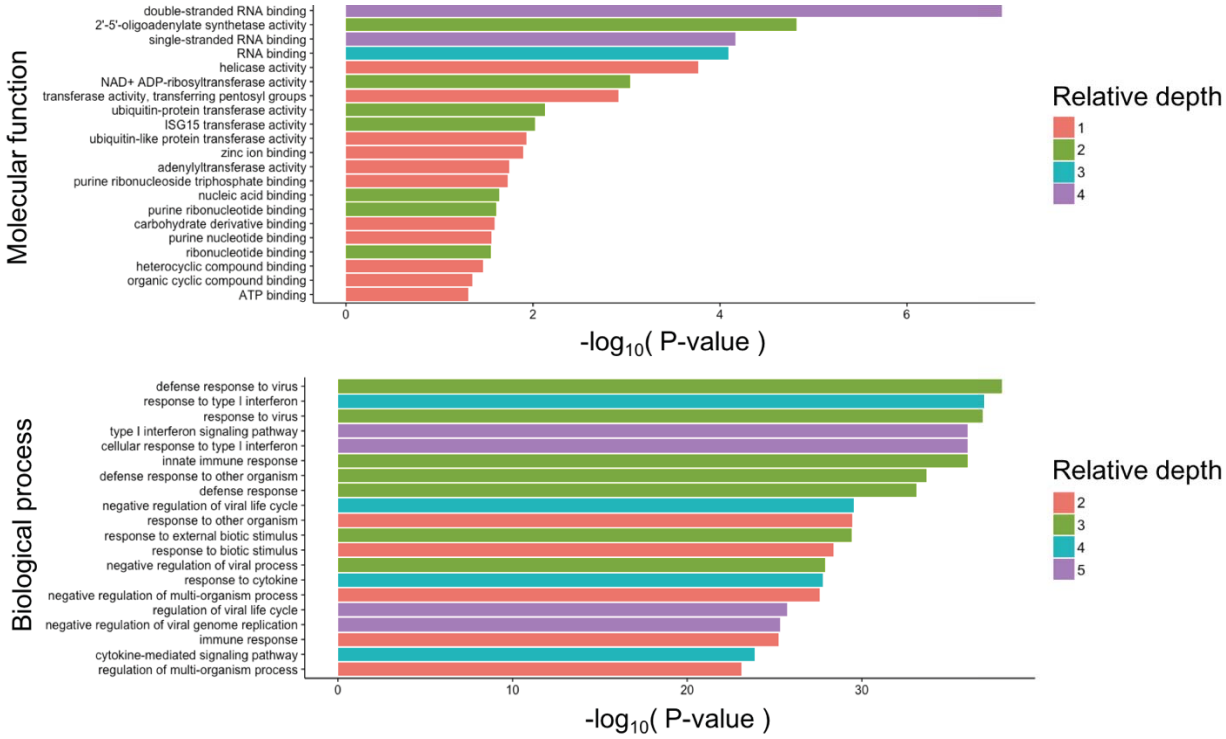

Figure 4.10 Polarisation of CD4$^+$ T cells with IFNβ activates the antiviral response The differentially upregulated genes upon polarisation with IFN-β were used to perform a GO term overrepresentation analysis using gProfileR. Results were ordered by adjusted P-value, with different colours representing relative depths within the ontology.

To determine whether the genes which contributed to the enrichment corresponded to functional blocks at the protein level, we built a protein interaction network using 133 upregulated genes as an input for STRING (169). Only interactions from experimental, co-expression and neighbourhood data were kept in the network. We observed a block of tightly interconnected proteins composed of genes from the OAS, and MX families (**Figure 4.11**).

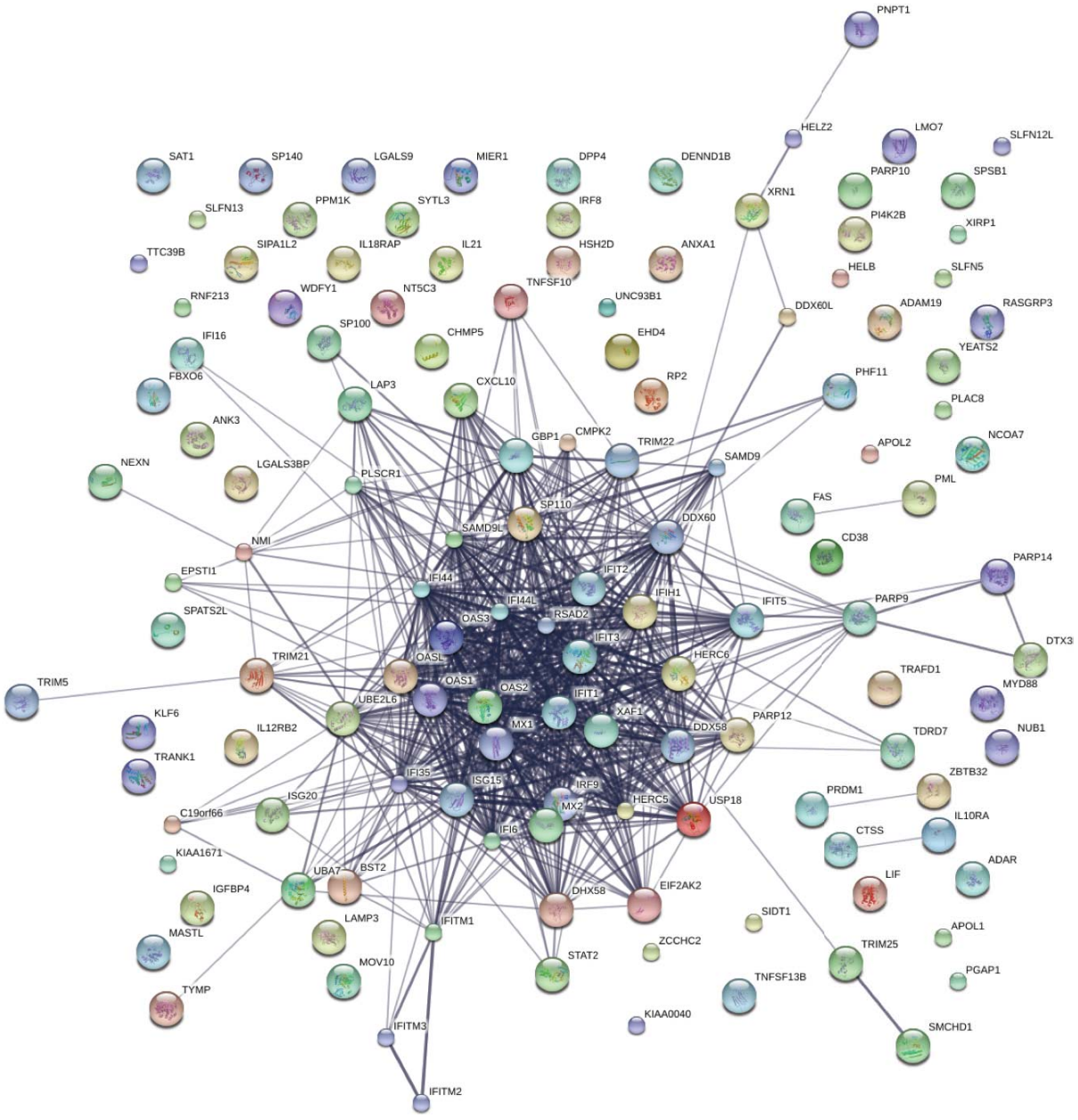

Figure 4.11 Protein network in CD4$^+$ T cells polarised with IFNβ Genes driving the enrichment in GO terms for antiviral response upon IFNβ stimulation were used to build STRING network. Connections represent protein interactions and co-expression expression. The width of each line is proportional to the strength of evidence for the interaction.

Then, we performed the same analysis for genes differentially upregulated in $CD4^+$ T cells. We used the results from T cell polarisation states at five days and observed that the genes upregulated in Th1 cells were enriched in signalling and response to cytokines (**Figure 4.12**). Specifically, Th1 cells upregulated the receptors for IFN-γ, IL-1, IL-18, and IL-33, as well as integrins and laminins. When analysing the 192 differentially upregulated genes in iTreg cells, we also observed enrichment in signalling and cell communication pathways (**Figure 4.13**). Here, the most represented molecular functions were chemokine receptors (CCRs), cytokine receptors and ion channels. Furthermore, we found that the genes upregulated in Th17 cells localised to integrin complexes and had membrane receptor or signal transducer activity. However, the upregulation of genes in Th2 cells seemed to be unspecific, with few GO terms showing only nominally significant P-values.

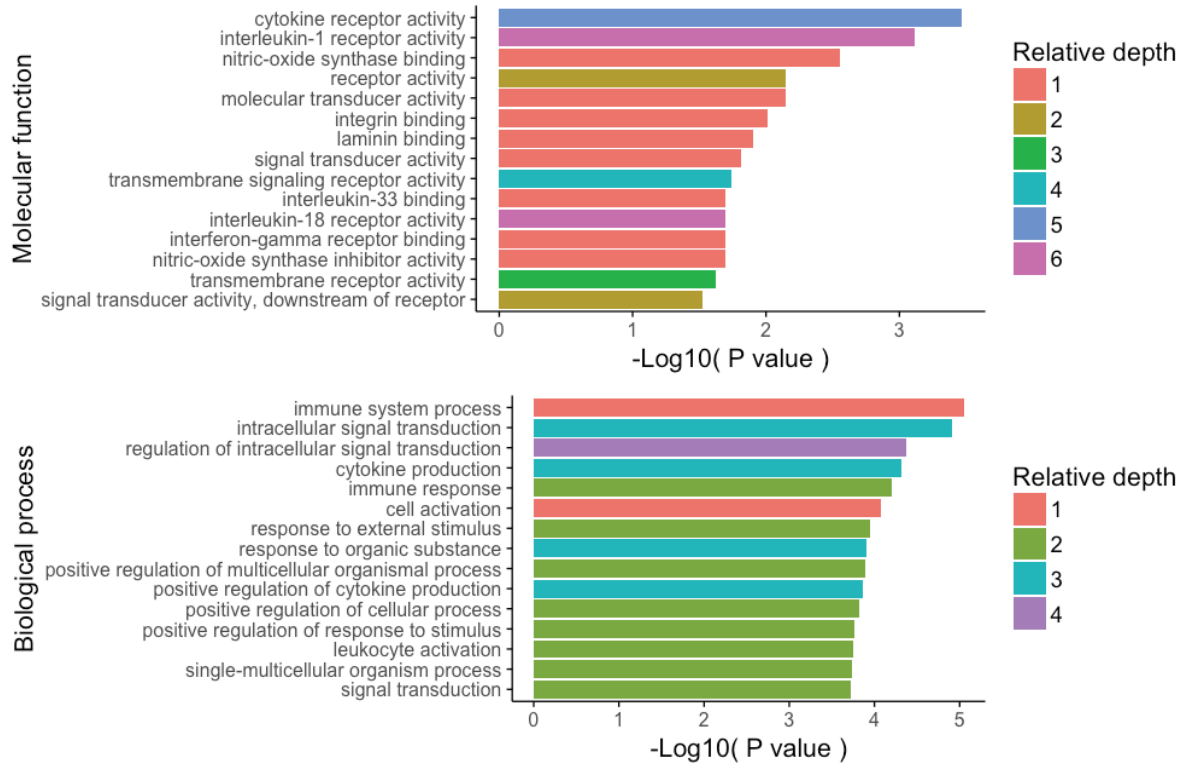

Figure 4.12 Genes involved in IFN and IL1 signalling are upregulated in Th1 cells Genes differentially upregulated upon polarisation to the Th1 lineage were used to perform a GO term overrepresentation analysis in gProfileR. The results were ordered by their adjusted P value, with different colours representing relative depths within the ontology. X axis represents the significance of the enrichment of molecular function and biological processes terms.

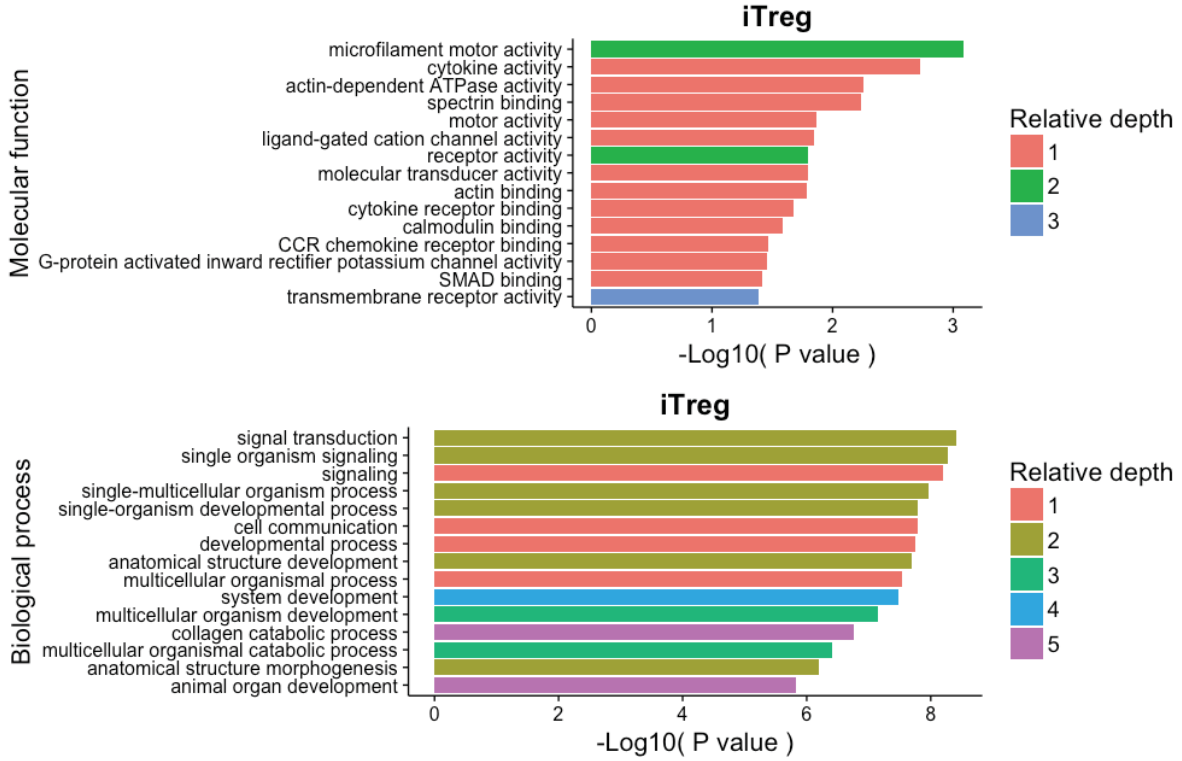

Figure 4.13 Genes upregulated by iTreg cells are enriched in chemokine and cytokine receptors Genes differentially upregulated upon polarisation to the iTreg lineage were used to perform a GO term overrepresentation analysis in gProfileR. The results were ordered by their adjusted P-values, with different colors representing relative depths within the ontology. Enrichment in molecular function and biological process terms in shown.

Next, we repeated the previous analysis for genes downregulated in $CD4^+$ T cells at five days. In Th1 cells we detected only 10 downregulated genes, therefore, we excluded this condition from the analysis. We observed that the 199 genes downregulated in Th2 cells were highly enriched in response to viruses, type I interferons (enrichment in the IFN response reactome pathway was also significant), and cytokines, with adjusted P values smaller than 1×10^{-20} (**Figure 4.14**). We also observed significant enrichment of IL-1 production pathways. The molecular function of these genes was annotated to families such as OAS, RNA-binding proteins and signal transducer terms.

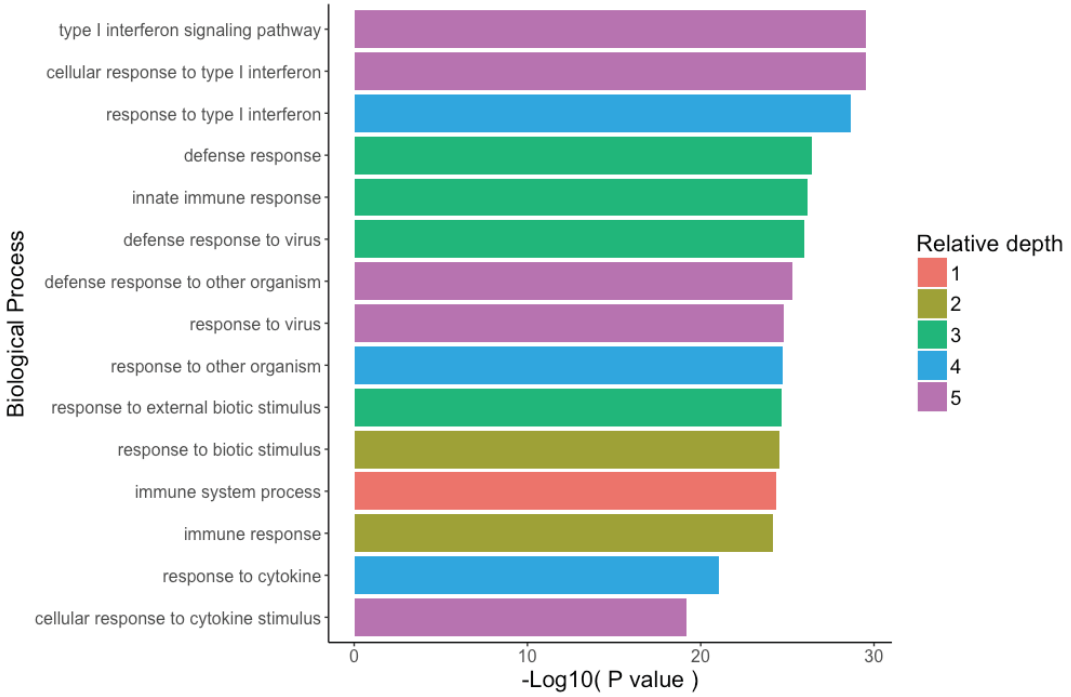

Figure 4.14 Genes involved in IFN signalling are downregulated in Th2 cells Genes differentially downregulated upon polarisation to the Th2 lineage were used to perform a GO term overrepresentation analysis in gProfileR. Results were ordered by their adjusted P-values, with different colours representing relative depths within the ontology.

Lastly, we analysed the downregulated genes in Th17 and iTreg cells. Th17 cells seemed to downregulate cell surface components such as integral membrane proteins. Interestingly, some of these genes localised specifically to phosphatidylinositol-3-kinase (PI3K) complexes. Furthermore, we also observed an enrichment in tyrosine kinase and PI3K activity (**Figure 4.15A**) and obtained similar results in iTreg cells (**4.15B**). Consequently, it is possible that downregulation of the PI3K pathway plays a role in Th17 and iTreg differentiation.

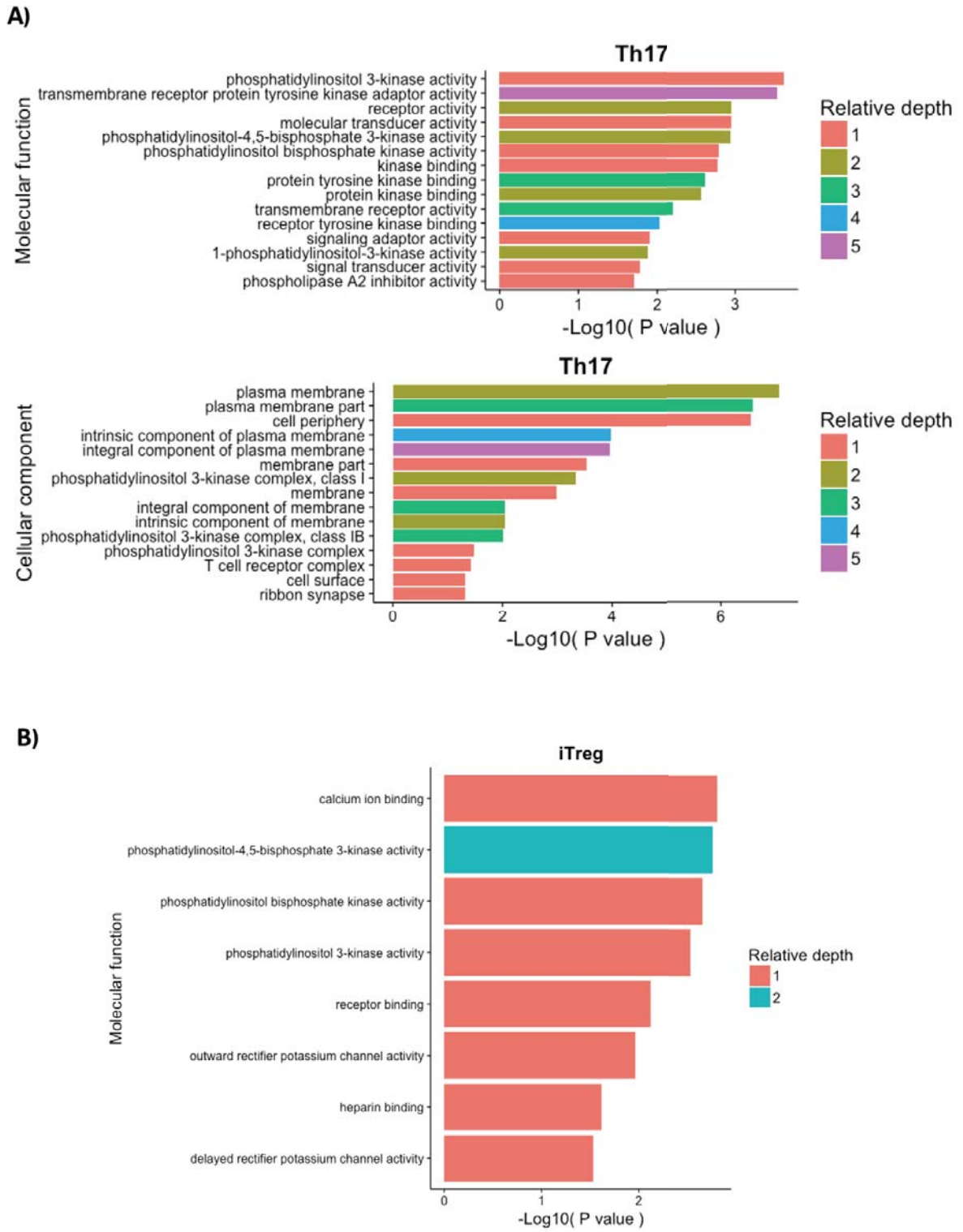

Figure 4.15 Genes of the PI3K pathway are downregulated in Th17 and iTreg cells Genes differentially upregulated upon polarisation to the Th17 and iTreg lineages were used to perform a GO term overrepresentation analysis in gProfileR. The results are ordered by adjusted P-values, with different colours representing relative depths within the ontology. **A)** Cellular component and molecular function term enrichment in Th17 cells. **B)** Molecular function term enrichment in iTreg cells.

4.6 Identification of potential lineage specific surface markers

Specific surface markers are necessary to study the role of cytokine induced polarisation *in vivo*. To address this issue, we tried to identify potential markers to isolate $CD4^+$ T cell subsets from the blood. We performed GO term overrepresentation analysis on the differentially upregulated genes which appeared only in one T cell state, but not in the others. Next, we examined the expression levels of the genes enriched in cell membrane categories. We identified five membrane receptors upregulated in Th2 cells: CD86, FZD3, ADGRA3, and the receptors for TGFβ and IL17 (TGFBR3 and L17RB) (**Figure 4.16**). We also observed TLR1 was differentially upregulated by Th1 cells (**Figure 4.17A**). Furthermore, four membrane receptors were upregulated only by Th17 cells: Lyn, TLR2, IL13RA, and integrin-β4 (IGB4) (**Figure 4.17B**). Finally, we identified three surface genes only present in iTregs: CD82, CD83, and CD101 (**Figure 4.17C**).

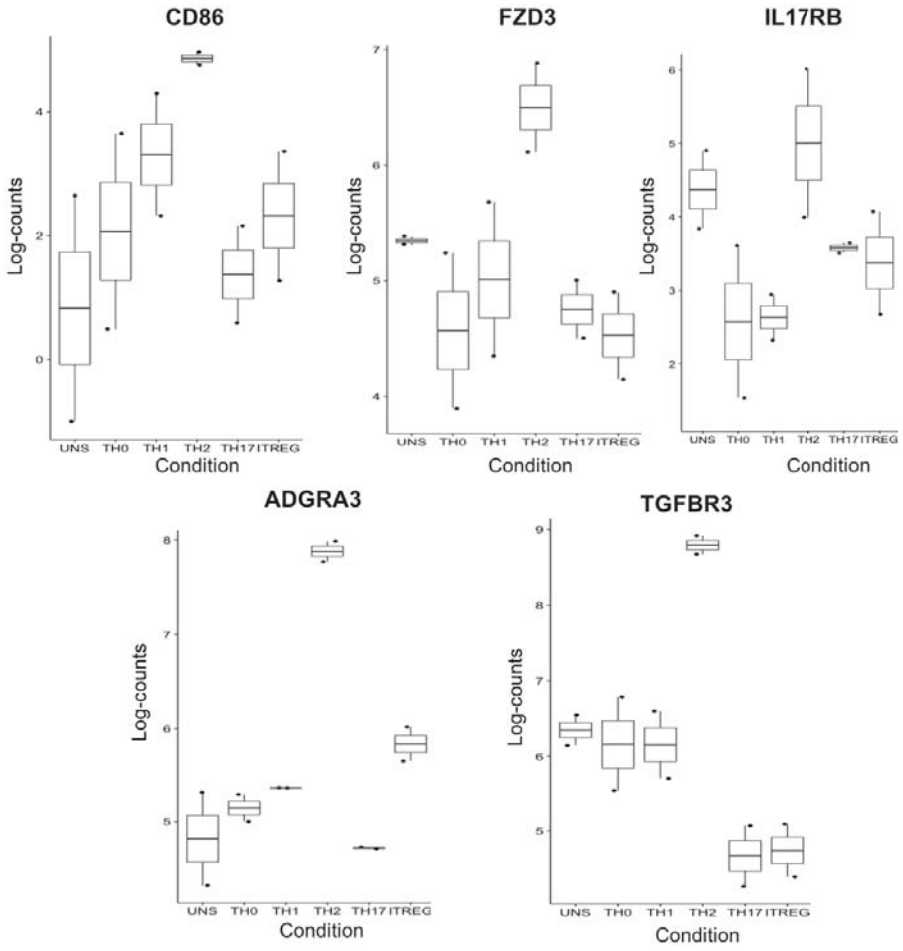

Figure 4.16 Differential expression analysis identifies potential surface markers for Th2 cells Upregulated gene specific to Th2 cells were used to perform a GO term overrepresentation analysis in gProfileR. The regularised log2 (rlog) counts for genes in cellular component category “membrane receptors” were inspected using bar plots.

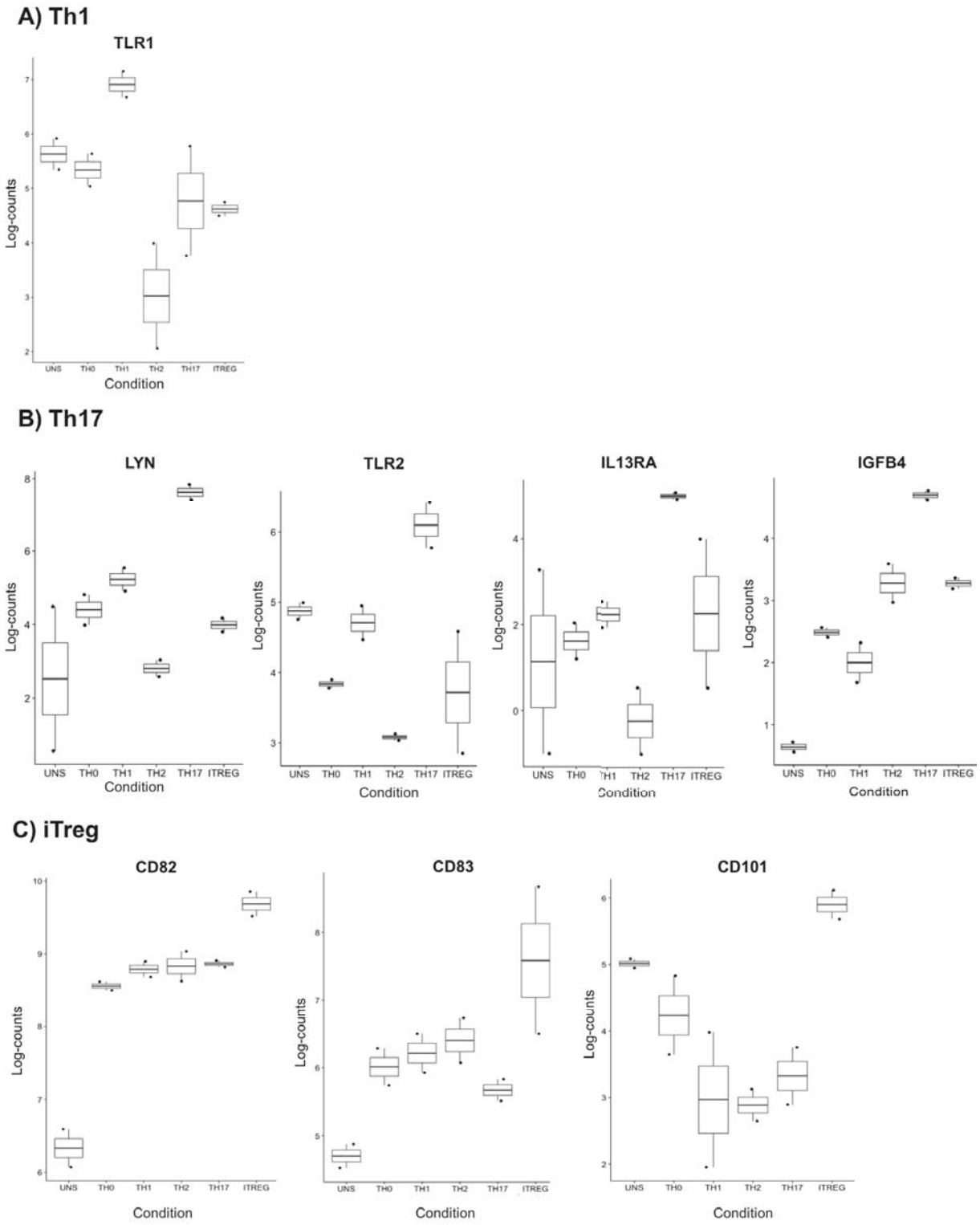

Figure 4.17 Differential expression analysis identifies potential surface markers for Th1, Th17 and iTreg cells Upregulated genes specific to each T cell subset were used to perform a GO term overrepresentation analysis in gProfileR. The regularised log2 (rlog) counts of genes in the cellular component category “membrane receptors” were inspected using bar plots. **A)** Th1, **B)** Th17, and **C)** iTreg cells.

4.7 Gene co-expression network analysis

We asked whether functional gene blocks were associated to specific transcriptional regulation mechanisms. For this, we performed a co-expression network analysis using the RNA counts from $CD4^+$ T cells stimulated for five days. At first, we focused on the type I interferon signalling pathway, which was downregulated in Th2 differentiation. We selected the interferon induced protein 35 (IFI35) gene, since it was present repeatedly in all the enriched GO terms related to IFN signalling. We calculated IFI35's co-expression network and annotated the genes in the network using the TRANSFAC tool for TF binding prediction (149, 170) Finally, we performed an overrepresentation analysis for TFBSs. We observed a significant enrichment in binding sites for the interferon response factors IRF9 (ISGF3), IRF8 (ICSBP), IRF1 and IRF7 (**Figure 4.18A**). This suggested that IRFs could be downregulated in Th2 cells, subsequently not binding to its targets and suppressing the IFN response. We confirmed that IRF8 and IRF9 expression was lower in Th2 cells than in other conditions (**Figure 4.18B**).

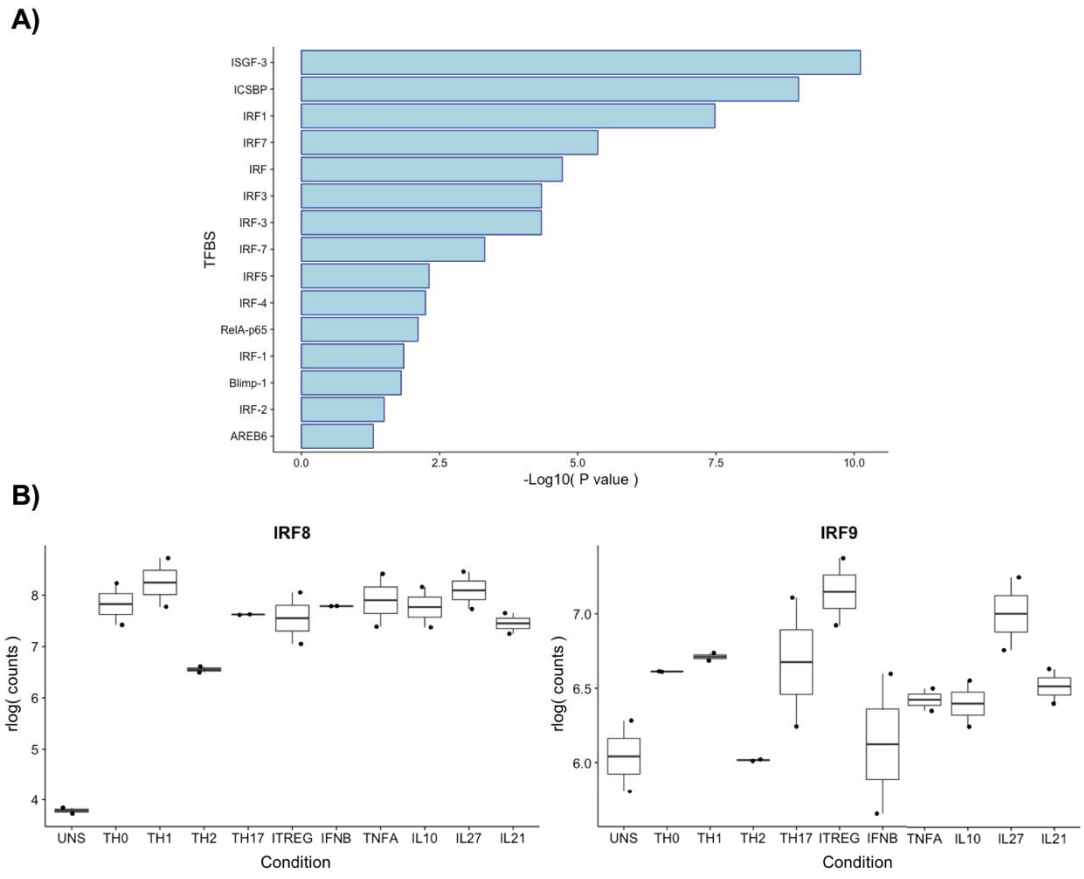

Figure 4.18 Downregulation of IRFs may suppress type I interferon responses in Th2 cells The gene IFI35 was used to build a co-expression network based on gene wise Pearson correlations. The members of this network were later used in a TFBS overrepresentation analysis with TRANSFAC in gProfileR. **A)** Enriched TFBS ordered by the adjusted P-values. **B)** Regularised log2 (rlog) RNA counts for IRF8 and IRF9.

Next, we applied the same approach to study the PI3K pathway, downregulated in Th17 and iTreg cells. To do so, PI3KR1 was used to build a co-expression network, since it was present in all the enriched terms related to PI3K activity. We evaluated the genes in this network using TRANSFAC and found a significant enrichment in TFBS for Sox4 and Msx2 (**Figure 4.19**).

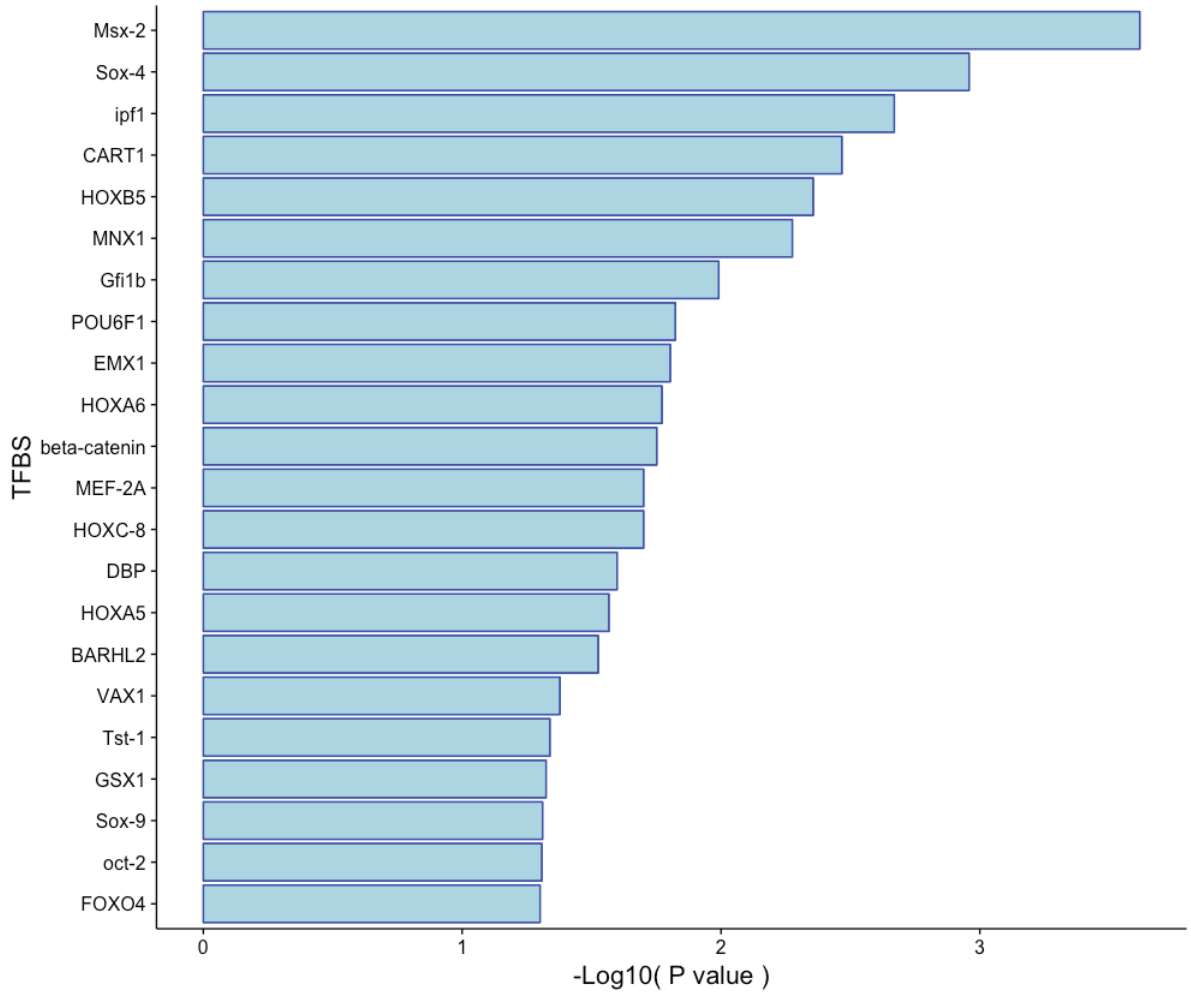

Figure 4.19 Downregulated genes in Th17 and iTreg cells are enriched in binding sites for Msx-2 and Sox-4 The gene PI3KR1 was used to build a co-expression network based on gene-wise Pearson correlations. The members of this network were later used to perform a TFBS overrepresentation analysis with TRANSFAC in gProfileR. The genes were ordered by the adjusted P-values.

4.8 Differential gene expression analysis in polarised macrophages

We were interested in functionally characterising the response of macrophages to cytokines. Consequently, we performed differential expression analysis to compare the different conditions with unstimulated macrophages. Approximately 690 genes were differentially expressed in M1 macrophages, and 500 in macrophages polarised with TNF-α. Moreover, we observed more than 200 differentially expressed genes in both M2 and IL-26 polarisation. On the contrary, stimulation with IL-23 caused no observable changes (**Figure 4.20A**). Detailed information concerning differential expression analysis of these conditions, such as fold changes, P-values and mean gene counts, is presented in the end of this thesis (**Appendix**).

Next, we asked whether transcriptional responses to cytokines were condition specific. We intersected the lists of differentially expressed genes of each condition and observed genes which seemed to only be present upon M1, M2, and TNF-α induced polarisation (**Figure 4.20B**). On the contrary, more than 50% of the genes upregulated in the IL-26 condition were shared with TNF-α (**Figure 4.20B**). We concluded that macrophages acquired different functions in response to different cytokines, and hypothesised that IL-26 activated similar pathways than TNF-α.

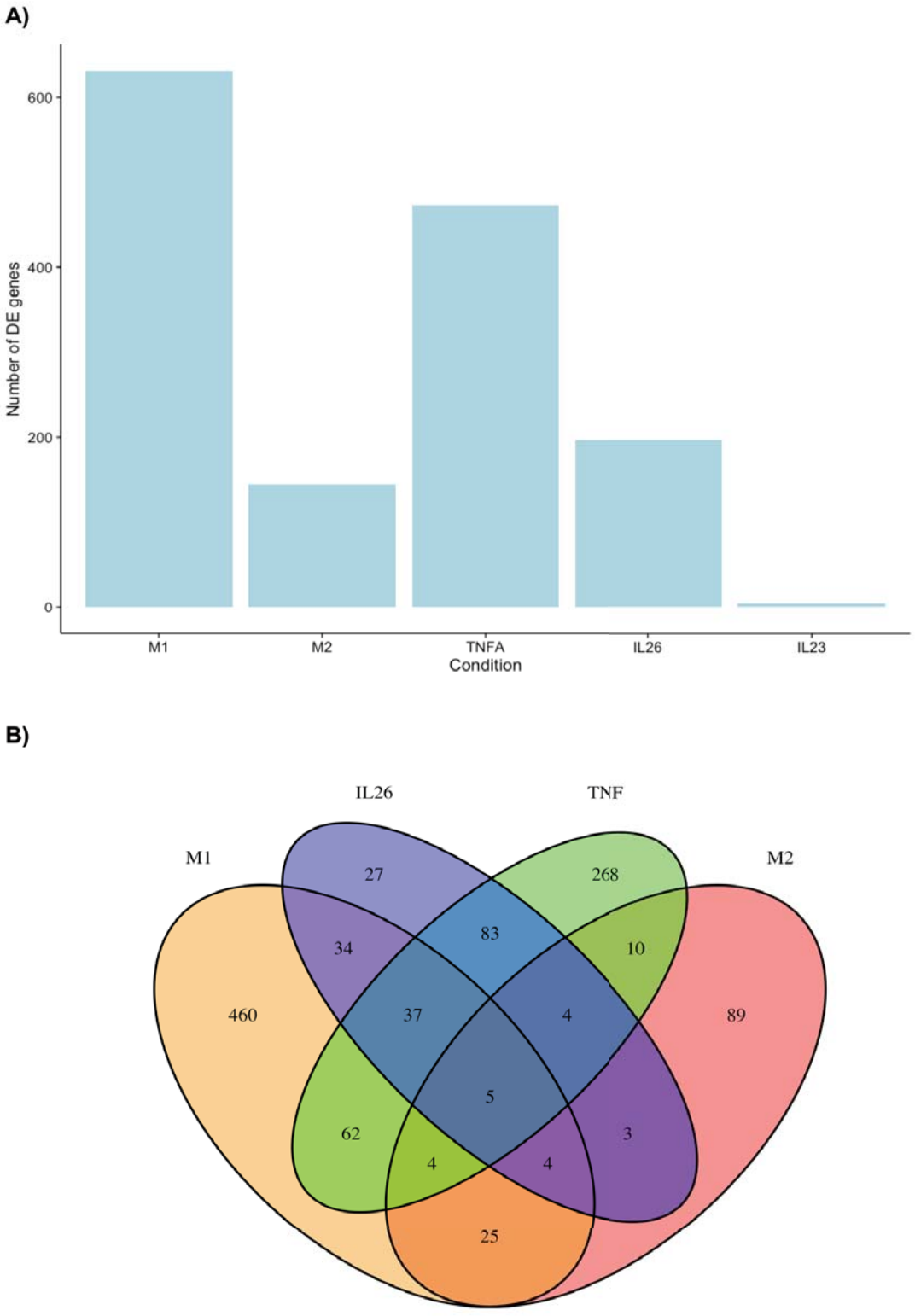

Figure 4.20 Number of differentially expressed genes upon cytokine induced macrophage polarisation Differentially expressed genes were defined by comparing macrophages polarised with cytokines to unstimulated macrophages. Genes with an adjusted P value ≤ 0.05 and an absolute $\log_2$ fold change ≥ 1 were considered differentially expressed. **A)** The numbers of differentially expressed genes after six hours were plotted. **B)** Overlap of differentially expressed genes across macrophage polarisation states at six hours.

To verify this, we performed GO term annotation followed by overrepresentation analysis in TNF-α and IL-26 stimulated macrophages. Both conditions were enriched in binding sites for NFκB (**Figure 4.21**). However, some of the molecular functions enriched in TNF-α were not present in IL-26 polarisation. For example, activation of the death domain and the PI3K pathway (**Figure 4.22**). We concluded that macrophage responses to TNF and IL-26 share a common regulation involving NFκB, but that TNF generated a broader response.

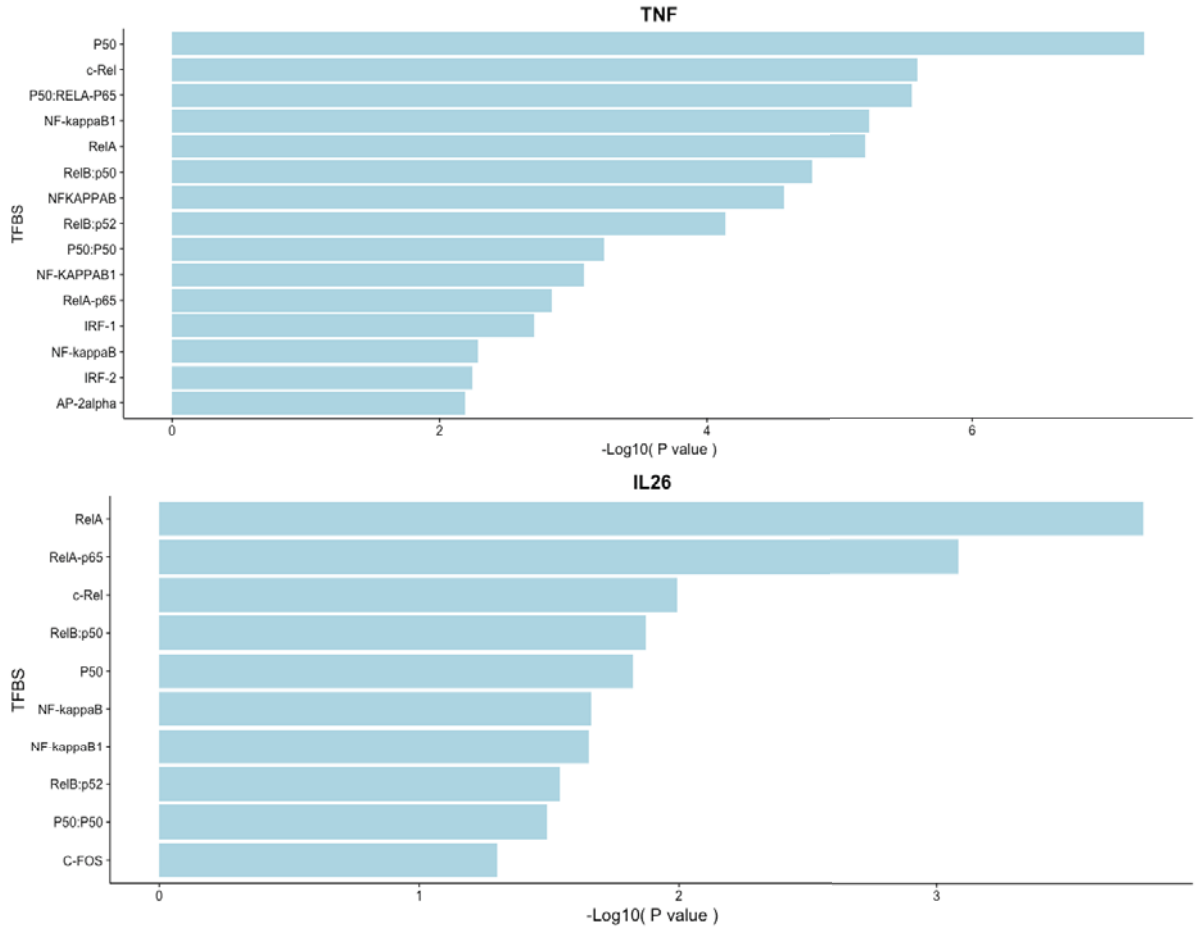

Figure 4.21 The responses to TNF and IL26 in macrophages are regulated by NFκB Genes differentially upon macrophage polarisation with TNF-α and IL26 were used to perform a TFBS overrepresentation analysis in gProfileR and TRANSFAC. Results were ordered by adjusted P-values.

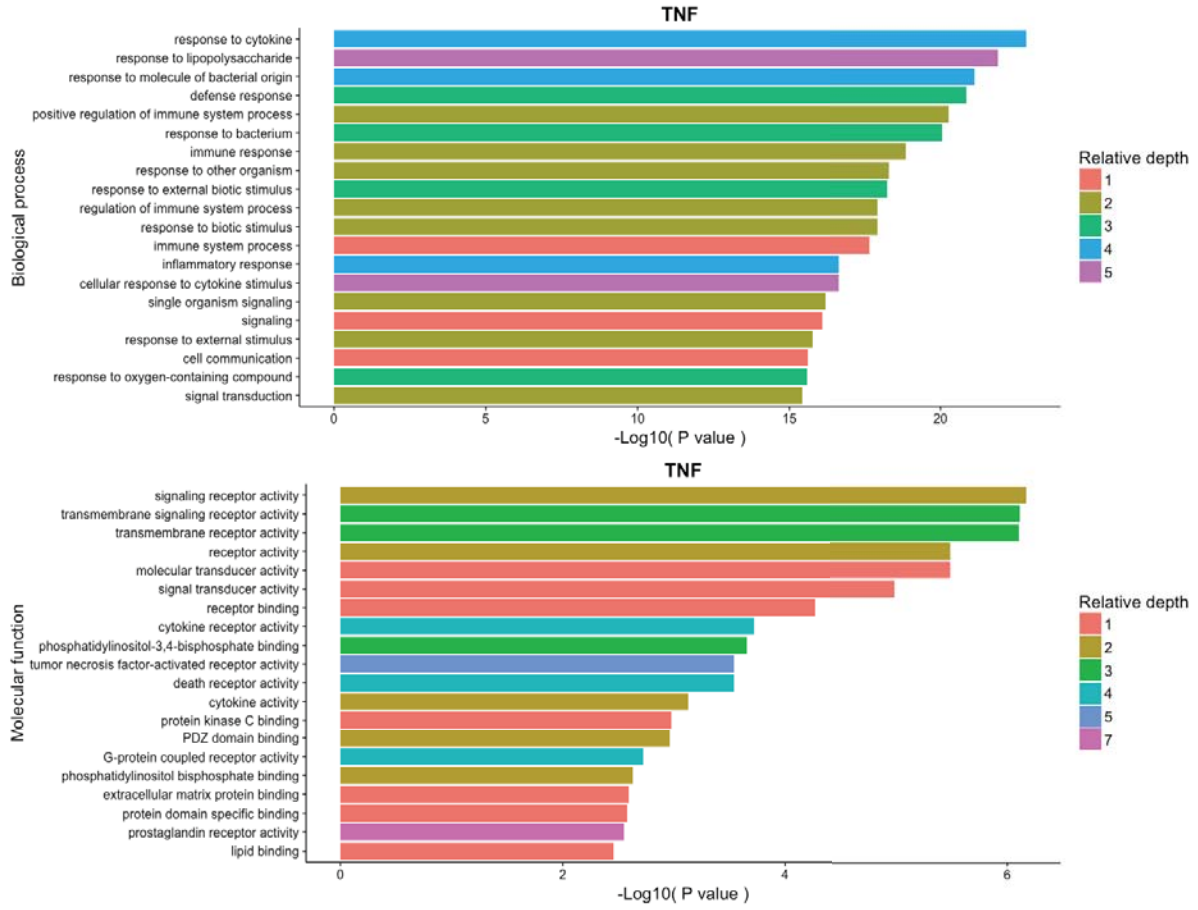

Figure 4.22 Macrophages upregulate proinflammatory in response to TNF-α Genes differentially upregulated upon macrophage polarisation with TNF-α were used to perform a GO term overrepresentation analysis in gProfileR. The results were ordered by adjusted P-values, with colours representing different depths within the ontology.

Finally, we characterised the effects of IFN-γ (M1) and IL-4 (M2) on the transcriptome of macrophages using GO term overrepresentation analysis. We found that genes upregulated in M1 macrophages were enriched in functions such as synthesis of OAS, ribonucleotide binding, cysteine peptidases, and other antiviral response pathways. (**Figure 4.23**). Consequently, we confirmed that macrophage polarisation with IFN-γ generated a proinflammatory phenotype tailored to respond against viral infections. On the other hand, half of the genes differentially expressed in M2 macrophages were downregulated. These genes were enriched in cytokine production pathways (**Figure 4.24**). Specifically, we found significant downregulation of IL-6 and IL-1β synthesis. Furthermore, we observed that genes upregulated in M1 macrophages are enriched in TFBS for STAT1, while those upregulated in M2 macrophages are enriched in TFBS for STAT6 (**Figure 4.25**). We concluded that M2 macrophages do not display an inflammatory phenotype, in contrast to their M1 counterpart, and that each polarisation acts via a different STAT.

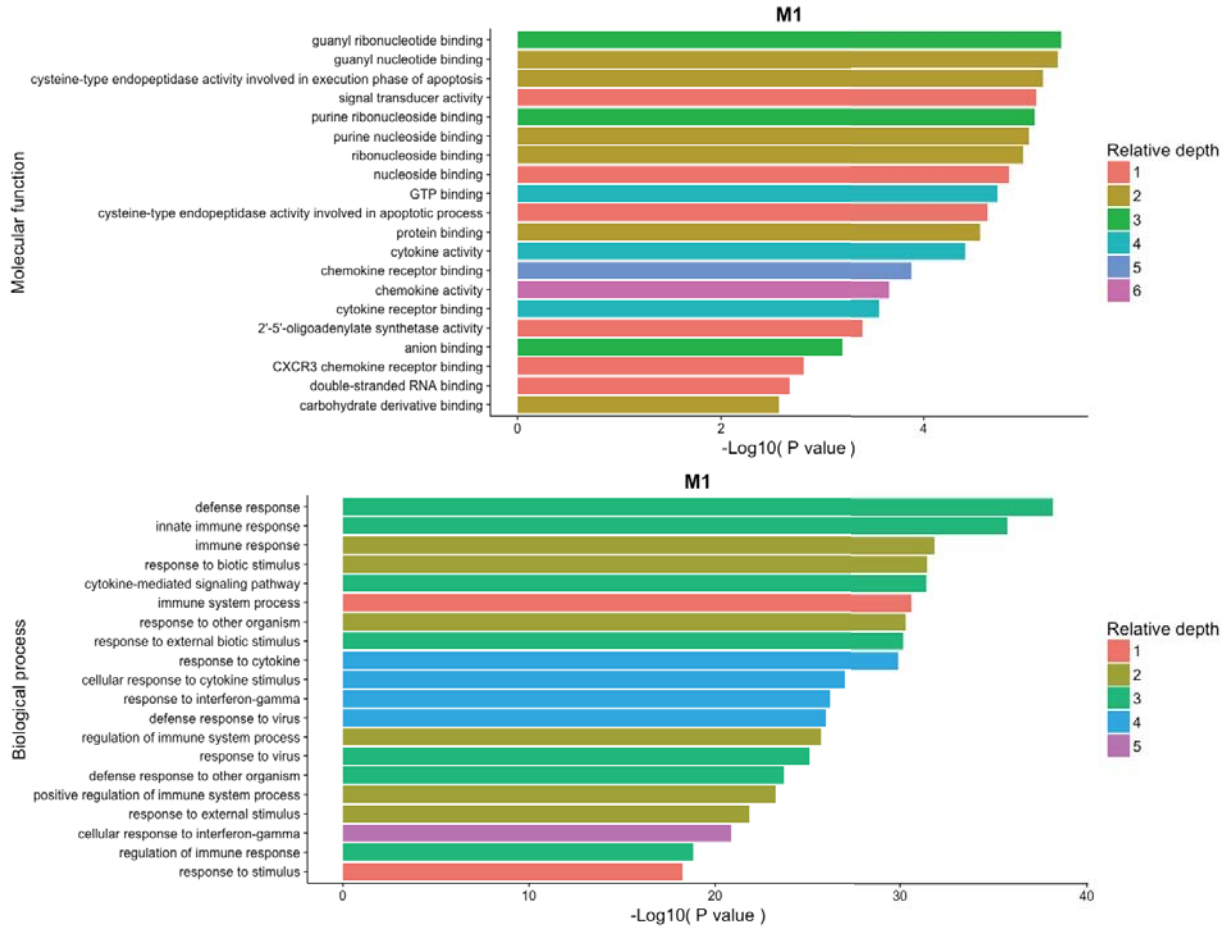

Figure 4.23 Macrophages activate the antiviral response upon M1 differentiation Genes differentially upregulated upon macrophage polarisation with IFN-γ were used to perform a GO term overrepresentation analysis in gProfileR. Results were ordered by adjusted P-values, with colours representing different depths within the ontology.

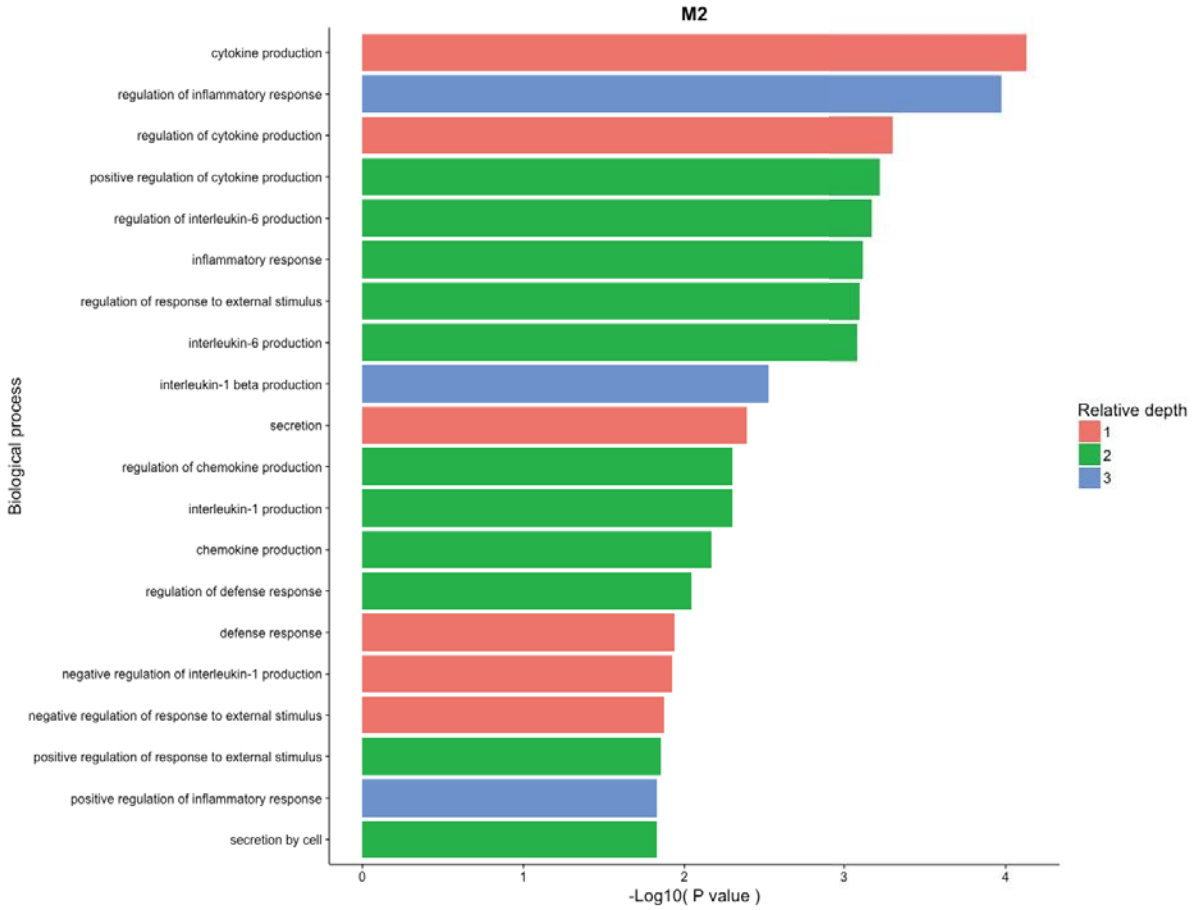

Figure 4.24 Macrophages downregulate the production of IL6 and IL1β upon M2 differentiation Genes differentially downregulated upon macrophage polarisation with IL-4 were used to perform a GO term overrepresentation analysis in gProfileR. Results were ordered by adjusted P value, with colours representing different depths within the ontology.

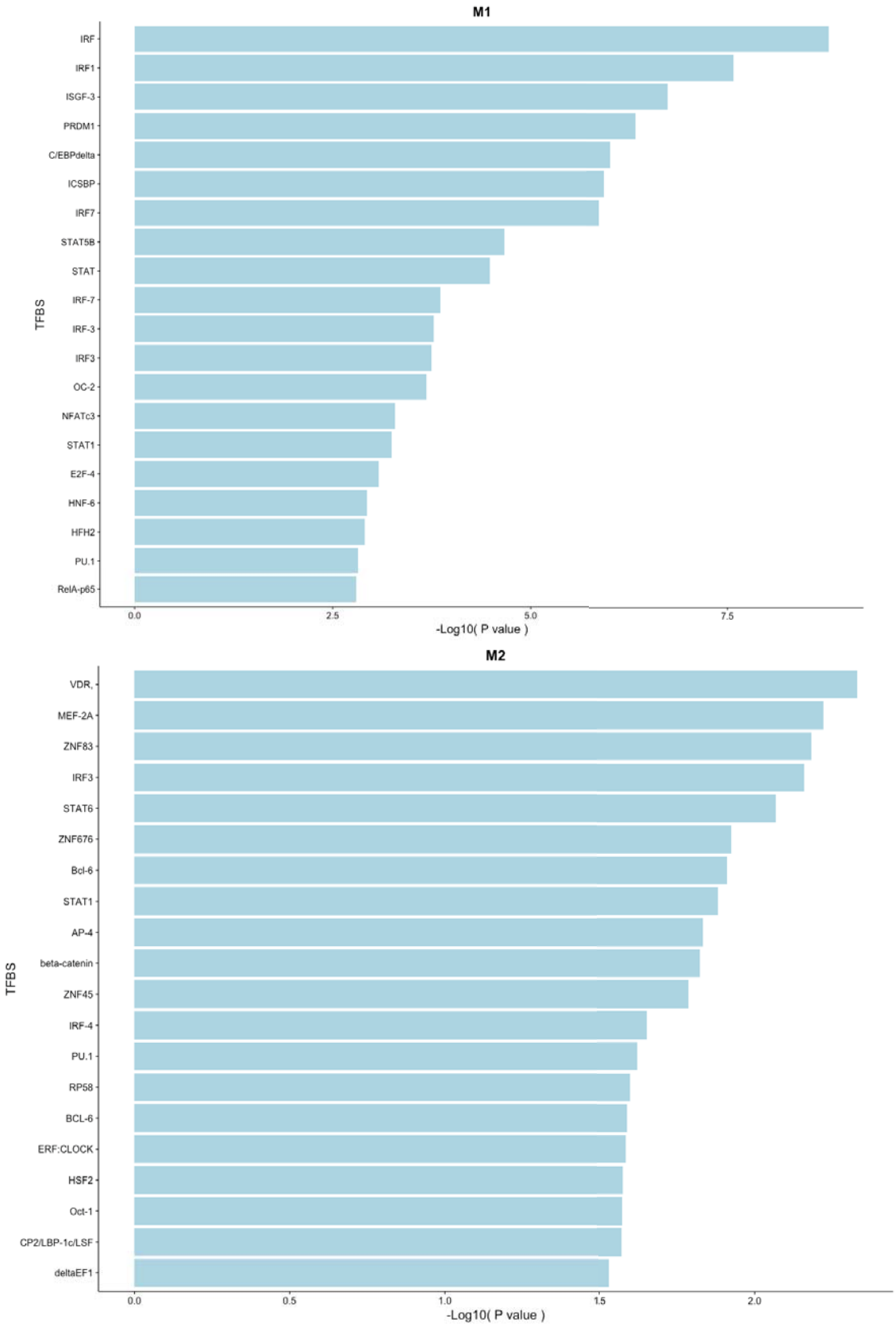

Figure 4.25 M1 and M2 macrophages activate signalling via STAT1 and STAT6 Genes differentially upregulated in M1 and M2 macrophages were used to perform a TFBS overrepresentation in TRANSFAC. Results were ordered by P-values.

4.9 Analysis of protein expression in stimulated CD4$^+$ T cells

Next, we asked if the results from RNA-seq of CD4$^+$ T cells could be reproduced at the protein level. We isolated the whole proteome of unstimulated CD4$^+$ T cells, Th0 cells and Th1 cells, and performed LC-MS/MS for protein identification and quantification. In order to assess the reproducibility of the technique, we conducted the experiment using two technical replicates of each condition. When comparing the raw counts from both technical replicates, we obtained a correlation coefficient of 0.99 across all conditions (**Figure 4.26**). Thus, we concluded that LC-MS/MS with isobaric tagging is a highly reproducible technique for protein quantification. Next, we performed data normalisation and PCA. We saw a clear separation by polarising condition, with the technical replicates clustering tightly together (**Figure 4.26**). PC1 separated the unstimulated samples from the remaining conditions and accounted for 71% of the observed variance. Conversely, PC2 explained separated Th0 and Th1 cells and explained 15% of the variance. We concluded that the technique was suitable for evaluating protein expression in CD4$^+$ T cells.

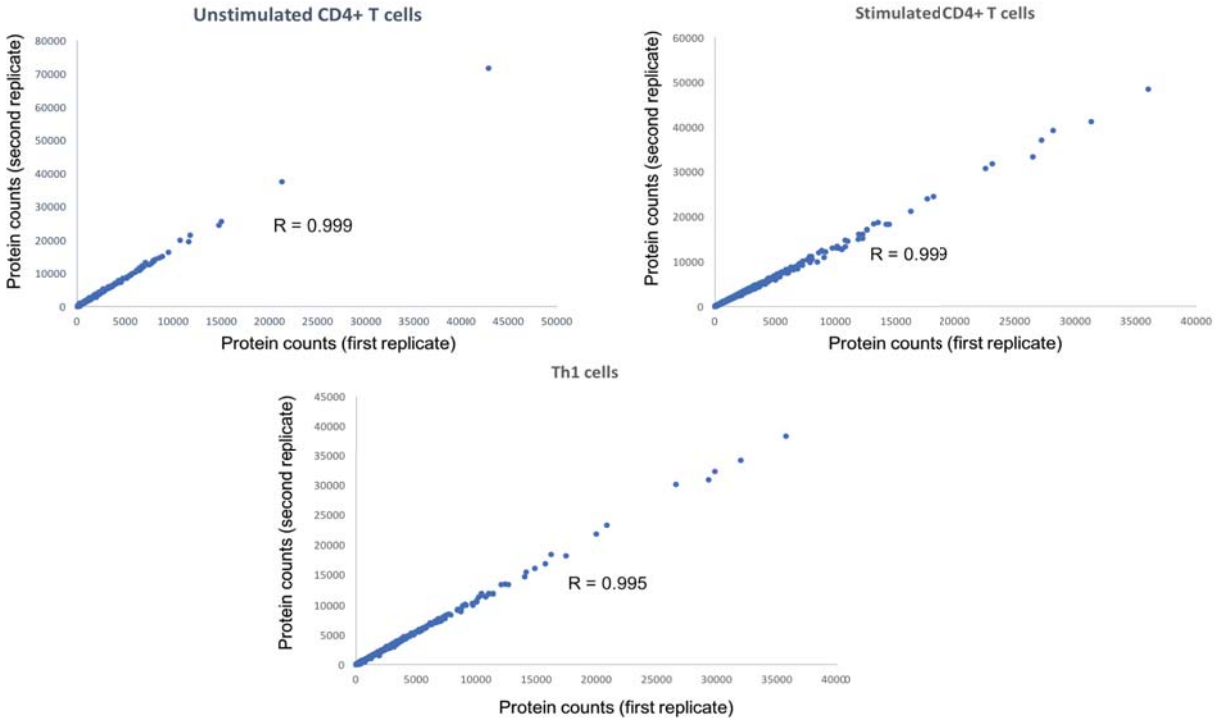

Figure 4.26 LC-MS/MS reproducibly quantifies protein expression in CD4$^+$ T cells Proteins from unstimulated CD4$^+$ T cells, stimulated CD4$^+$ T cells and Th1 cells were quantified with TMT-labelling and LC-MS/MS. The results from two technical replicates were compared using scatter plots. The Pearson correlation coefficient of both replicates is shown in each plot.

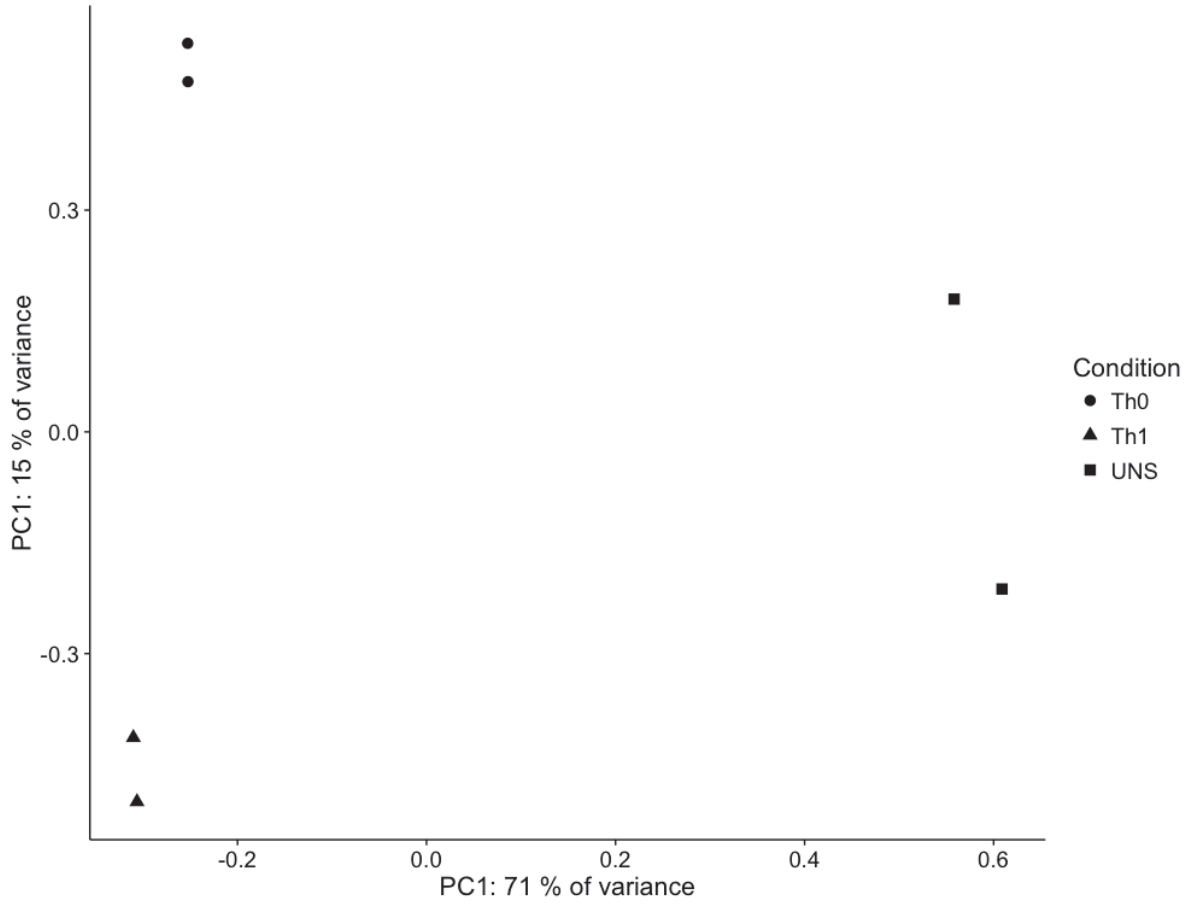

Figure 4.27 PCA separates stimulatory conditions in LC-MS/MS data Protein counts were normalized to the sample median and used to perform PCA. Different shapes represent different polarising conditions.

To quantify the similarity between RNA-seq and LC-MS/MS results, we computed the correlation of both data sets. To do this, we calculated the $\log_2$ fold changes between unstimulated cells and Th0 or Th1 cells at both the protein and RNA level. Next, we computed the sample wise Pearson correlation between both methods. We concluded that the changes in mRNA and protein levels were correlated (**Figure 4.27**), with Pearson coefficients close to 0.5. Through visual inspection, we identified genes with particularly high correlations between RNA-seq and LC-MS/MS (**Figure 4.27**). These genes were enriched in cell cycle functions. Based on these results, we concluded that RNA and protein changes in $CD4^+$ T cells were correlated and reproducible, and that both techniques could be used to study cytokine induced cell polarisation.

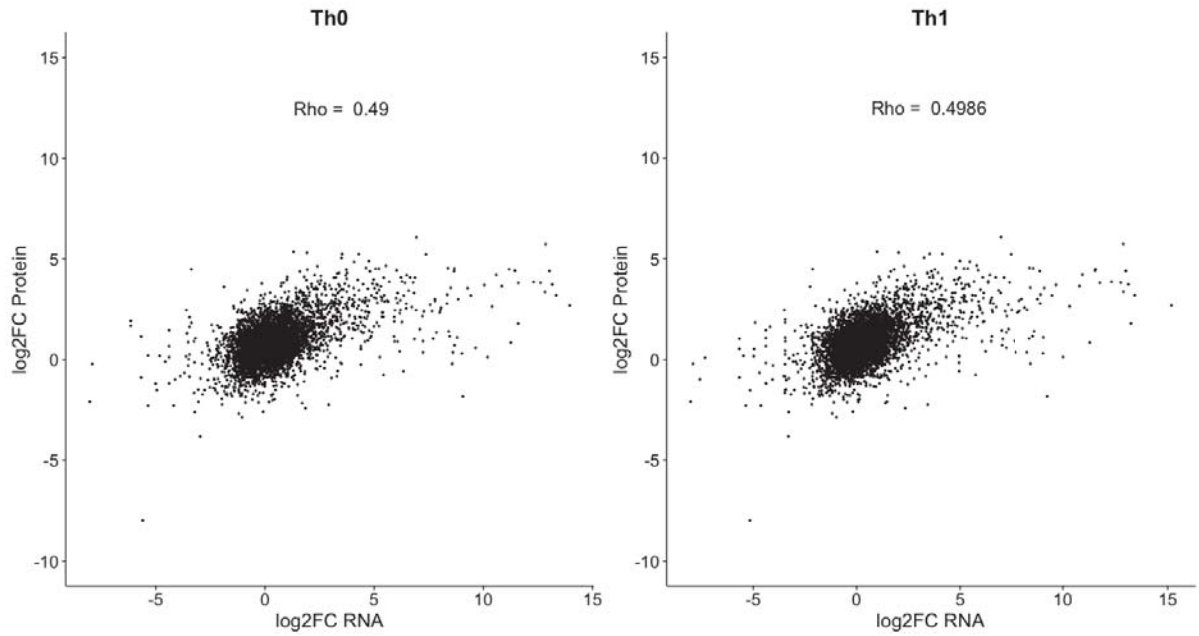

Figure 4.28 RNA-seq and LC-MS/MS data are highly correlated Log2 fold changes between Th0/Th1 and unstimulated $CD4^+$ T cells were calculated using LC-MS/MS and RNA-seq data. Next, a scatter plot of both data sets was built. The Pearson correlation coefficient between both methods is shown in each plot.

4.10 Discussion

In this chapter, I showed a characterisation of the transcriptional response to cytokines in human macrophages and T cells. This characterisation was performed using low coverage RNA-seq, an affordable approach to examine a large number of conditions (171). At first, I summarised the most important sequencing quality metrics. These results suggested that the RNA-seq run was not only performed adequately, but also generated reproducible data.

We acknowledged that a variety of power limitations might be present in this study. This is due to the fact that, to evaluate a large number of conditions with a limited budget, sequencing depth was kept at $5x10^6$ PE reads and the sample size was only two replicates. Since it is known that statistical power depends on the effect size, coverage and sample size (132-134), this raises potential issues. Hence, we modelled the power of the study to detect differences in genes with varying expression levels. We concluded that, if an average fold change of two is assumed, we might be powered to confidently detect differential expression in genes with raw counts higher than 50. Since the average count of genes in this study is 52, we concluded that our power is close to 0.8. Despite the small sample size, we achieve a relatively high power. This might be explained by the statistical model implemented in

DESeq2, specifically tailored for low sample size (135) and shown to achieve the highest power under these circumstances (132). Since our starting material are cells belonging to the same cell type and kept in cell culture, a tightly controlled environment, the coefficient of variation is also relatively low, which might contribute to the relatively high sensitivity (132). Nonetheless, there exist power limitations, particularly for lowly expressed genes. These should be acknowledged when evaluating aspects as condition specific gene expression or conditions with no observable differences. Each conclusion will be dependent on the mean expression level of the genes in question, which can be accessed via the Appendix to this thesis.

Having investigated our power limitations, we analysed differential gene expression. Initial inspection of the data revealed several combinations of cytokines and time points which seemed to show no effect. For example, we observed no differential gene expression after 16 hours of $CD4^+$ T cell polarisation to Th1, Th17 and iTreg. However, we did find large differences compared to Th0 after five days of stimulation. This suggests that Th1 and Th17 differentiation occurs in the later stages of the immune response. However, since some of the effect sizes observed at these early time points might be relatively small, this must be verified in a more powered study, with more biological replicates and a larger sequencing depth. On the other hand, we observed Th2 differentiation in early stages of activation, and it persisted during the course of five days. The opposite is true for IFN-β, which induced major effects in T cells at 16 hours but almost no effect after five days. Macrophages also showed prominent transcriptional responses to IFN-γ, IL-4, TNF-α, and IL-26 after six hours. On the other hand, polarisation of $CD4^+$ T cells with TNF-α, IL-10, IL-21 and IL-27, and polarisation of macrophages with IL-23 revealed no changes at any time point, but it remains unclear whether this is biologically true or a consequence of the low power of this study: we cannot completely rule out an effect of these cytokines given the small sample size and low coverage. For the purpose of the following discussions I only consider conditions which showed differential gene expression.

The response of $CD4^+$ T cells to IFN-β was strikingly different to that to other cytokines. Consequently, we hypothesised that transcriptional changes upon IFN-β stimulation were transitional, while polarisation to other cell subsets was stable. However, this observation should be confirmed by examining more time points following T cell activation. An analysis of genes upregulated upon stimulation with IFN-β revealed that most of these genes act at an early time point. This is expected, as type I interferons are involved in the cellular response to viruses (172), which must be triggered immediately upon infection in order to prevent exponential reproduction of the pathogen. Our functional enrichment analysis also supported

this explanation. It revealed a network of tightly connected genes composed of oligoadenylate synthetases (OAS), involved in the activation of RNAse L, and IRFs (167, 173). The difference between transcriptional dynamics in Th2 (the Th subset with the largest number of differentially expressed genes) and IFN-β could be explained by the differentiation model to lineages proposed in the literature in which initial activation of a master regulator TF induces expression of signature cytokines. These cytokines then induce the expression of the regulator TF in a self-sustaining positive feedback loop (25, 62, 174-176).

When analysing known T cell polarisation states, we observed that the cells obtained agreed with their descriptions in the literature. For example, Th1 and Th2 cells were originally described as $CD4^+$ T cells with distinct, mutually exclusive cytokine secretion profiles (14, 177). These observations have been extensively reproduced, with Th1 cells secreting IL-12 and IFN-γ, and Th2 cells producing IL-4, IL-10 and IL-13. These phenotypes are induced by the activation of the TFs T-bet and GATA-3, respectively (35, 36). All of these observations are reproduced in our data set. We observe a higher level of IFN-γ and T-bet mRNA in Th1 cells, and higher IL-4 and GATA3 transcription in Th2. The same is true for Th17 cells, which in our study express higher levels of IL17F and RORC at the mRNA level, as has been described elsewhere (25, 27, 37, 178). Furthermore, Treg cells are known to express the master regulator TF FoxP3 (179), which was also observed in our RNA-seq study of iTreg differentiation. It is worth noting that, in contrast to the common conception derived from mouse models that $CD4^+$ regulatory T cell function is dependent on IL10 and TGF-β (180-182), we only found slight upregulation of the IL-10 and unaltered expression of TGF-β. This, however, cannot be ruled out given our sample size and has to be verified in a study with larger power. Conversely, we saw upregulation of CTLA-4, which has been proposed as a key mechanism of Treg mediated immune regulation (182-185).

Macrophages polarised to the M1 phenotype are known to upregulate IL-12 and COX-1, while M2 macrophages express mannose receptors and COX-2 (83, 94, 95). We also recapitulated these findings. Taken together, our observations suggest that cytokine polarisation with this experimental approach is effective for both $CD4^+$ T cells and macrophages.

Next, we compared polarised $CD4^+$ T cells upon five days of stimulation against Th0 cells. We found only 45 differentially expressed genes between Th1 and Th0 cells, compared to over 250 in all other comparisons, which suggested that Th0 cells might polarise to a default functional state similar to Th1. This might be explained by the autocrine action of cytokines, since our cytometry data proved that TCR stimulation alone upregulates T-bet, which can

induce IFN-γ (35) and trigger a positive feedback loop (62). Furthermore, strong TCR stimulation could also explain this observation, since it is known that the strength and duration of TCR signalling can influence cell fate (33, 52, 186), with strong and prolonged TCR signalling generating Th1 cells (31). To disentangle the causes of this observation, we would need to perform another experiment using anti-IFN-γ antibodies to block the effects of autocrine cytokines.

When comparing the remaining lineages with Th0, we observed a combination of gene expression changes which appeared to be unique and non-overlapping for each condition. Given that the average power of our study is close to 80%, it is not possible to definitely state that these changes are condition specific. There is a possibility that these genes were indeed differentially expressed in other conditions to a smaller extent, and thus not identified by our statistical analysis. However, since the cytokine secretion profiles of Th1, Th2 and Th17 cells are known to be mutually exclusive (14, 16, 25, 187), we think that these observations might suggest the existence of transcriptional profiles specific to each condition. Next, we used the lists of differentially expressed genes only observed in each lineage and used them for surface marker discovery. This was motivated by the lack of markers for the isolation of specific $CD4^+$ T cell subsets *in vivo*, given that most of the surface receptors proposed in the literature have either not been sufficiently specific (54, 188), or been dependent on particular pathogenic stimuli (189). We identified a variety of potential markers, including TLR1 upregulated in Th1 cells, and TLR2 upregulated in Th17 cells. A comparatively large set of potential markers was obtained for Th2 cells, which included ADGRA3 and TGFBR3. Because marker discovery was solely based on RNA level and the study is not fully powered, further validation is needed. Our confidence on the specificity of each of these markers increases with their expression level, since we determined that the study is more powered to detect differential expression in highly expressed genes. In order to validate these markers, we would need to perform further experiments, using antibodies to stain the markers of interest and to purify populations enriched in them. Next, the cytokine secretion patterns of these populations would need to be characterised using Luminex or enzyme-linked immunosorbent assay (ELISA). We would expect, for example, cells enriched in markers identified here as Th2 specific to secrete IL-4, IL-10 and IL-13.

We also asked whether LC-MS/MS coupled with quantification by isobaric labelling is a reproducible technique for analysing protein expression in $CD4^+$ T cells. Our results indicated that this technique was not only highly reproducible, but also correlated well with the observations from RNA-seq in sample wise comparisons. These results suggested that

LC-MS/MS might be a powerful platform for surface marker discovery in $CD4^+$ T cells too, because it would allow unbiased screening of the cell's proteome. However, we would need to evaluate the efficiency of this technique for identifying membrane bound proteins.

Next, we focused on determining the functional context of differential gene upregulation and downregulation in each condition. We did this by gene ontology annotation followed by overrepresentation analysis of the GO terms. Some of the genes downregulated in Th2 cells were involved in the response to IFN-γ, the hallmark Th1 cytokine. Furthermore, the response to type I interferons was downregulated to an even larger extent. It is known that Th2 and Th17 differentiation are inhibited by both IFN-γ and IFN-α (25, 190). However, once differentiation has occurred Th17 cells are able to retain their phenotype even in the presence of IFN-α (25), while the Th2 phenotype can be reversed (190). Thus, these results point to the activity of IFNα/β signalling being a crucial, tightly regulated determinant of Th1/Th2 responses. When using co-expression network analysis to study the type I IFN enriched pathways, it became clear that most of the genes have binding sites for IRFs, specially IRF9 and IRF8. IRF9 is necessary for the response to IFN-α, IFN-β, and also IFN-γ (173, 191). On the other hand, IRF8 is specifically induced by IFN-γ, but not by type I interferons, and it interacts with PU.1 to drive T cell differentiation (173). This suggests that the response to IFN-γ is downregulated in Th2 cells, which might be seen as a suppression of the Th1 differentiation program. These observations seem to support the mutual exclusivity model of Th1/Th2/Th17 differentiation discussed above.

By performing a gene co-expression analysis of downregulated genes, we observed enrichment in binding sites for Sox4 in iTreg and Th17 cells. A Sox4 *knock in* study in $CD4^+$ T cells showed that this TF is directly induced by TGF-β and is involved in suppression of Th2 differentiation (192). As a part of that study, ChIP-seq for Sox4 in $CD4^+$ T cells was performed, and it was concluded that Sox4 suppresses Th2 differentiation by directly inhibiting GATA-3 binding (192). Furthermore, Sox4 also modifies the chromatin landscape by regulating the expression of histone deacetylases in T cells (193). This evidence suggests multiple functional roles for this protein. Results from our RNA-seq analysis contribute to this discussion by suggesting a potentially new functional mechanism: direct downregulation of Th2 related genes by Sox4. However, this hypothesis must be validated by functional studies capable of revealing TFBS occupancy in Sox4 targets upon cytokine induced polarisation and no definitive conclusions can be reached so far.

When analysing gene expression in iTreg cells, we found that iTregs share a large proportion of their transcriptional program with Th17 cells. This is expected, as both cell

types have been previously associated in the literature. For example, iTreg cells can convert to Th17 cells in response to pathogenic signals (65, 178). We also identified several overlapping functions between these lineages. For example, downregulated genes are enriched in the PI3K pathway in both cell types. This pathway is involved in T cell polarisation (194), where PI3Ks phosphorylate PIP_2 and generate PIP_3, which in turn causes the activation of Akt, a central component of the mTORC pathway (195). Since Akt inhibits the function of FoxO1 and FoxO3a (195), regulators of FoxP3 (196), inhibiting the PI3K pathway is necessary for Treg differentiation (197). This agrees with our observations. However, downregulation of PI3Ks in Th17 cells is puzzling, since the PI3K-Akt-mTORC1 axis is thought to promote Th17 differentiation through translocation of RORγ to the nucleus (195, 198, 199). Given the fact that PI3K downregulation is highly enriched in Th17 cells in our data set, our observations seem to contradict this model. Functional validation of these observations is needed to clarify whether up or downregulation of this pathway is necessary for Th17 differentiation. The observations described above highlight crucial aspects of Th17 biology. They are of particular importance in the study of MS, which is largely driven by Th17 responses (108, 200, 201). Disentangling the contributions of these factors in $CD4^+$ T cell fate decision towards the Th17 lineage, and specially in the balance between Th17 and Treg cells, could aid in proposing new drug targets for MS treatment, which is unresponsive to widely used autoimmunity therapies such as TNF blockade (123).

Unlike MS, RA is usually responsive to TNF-α and IL-6 blockade (2, 115, 119). As a matter of fact, TNF-α is involved in multiple autoimmune diseases (119). This encouraged us to characterise the effects of TNF-α in different cell subsets. TNF-α is usually regarded as a cytokine of the innate immunity (2) which acts on macrophages and mononuclear phagocytes (119). Our results suggest that, indeed, stimulated $CD4^+$ T cells might not respond to TNF-α treatment. However, the power limitations of our study make it difficult to draw any definitive conclusions. Conversely, macrophages upregulated more than 400 genes in response to TNF-α. The response of macrophages to TNF-α was apparent at six hours, and a more detailed analysis revealed that upregulated genes were enriched in TFBS for NFκB and IRFs. These same genes were also enriched in death domain activity and receptors for cytokines and prostaglandins. The fact that both the death domain and NFκB are enriched suggests that macrophages might be responding by activating more than one of these pathways.

Furthermore, IL-26 has also been associated to autoimmunity (202). Interestingly, macrophages responded to IL-26 in a similar way as to TNF-α, with 83 out of 200 differentially expressed genes being shared between both conditions. The IL-26 response

was also enriched in genes controlled by NFκB, but the results from pathway enrichment analysis suggested a slightly different profile, which did not include activation of the death domain. This suggests that the overlap between the response to IL-26 and TNF-α might be due to the shared elements in their signalling cascade and not to TNF-α autocrine secretion in response to IL-26. However, to completely rule out the role of autocrine cytokines we would have to conduct an experiment in which macrophages were stimulated with IL-26 while supplementing the culture with antibodies against TNF-α. It is important to determine the source of this shared signalling in light of the currently existing TNF-blockade therapeutics. Given that anti-TNF therapy is not effective in a subset of patients, and having shown that TNF-α generates prominent effects in macrophages, it would be interesting to determine whether the patients who do not respond to therapy show any additional IL-26 signalling. If this is proven true, therapies focused on IL-26 blockade could be investigated.

IL-23 has also been linked with autoimmune reactions involving macrophages, particularly in the context of MS and inflammation of the central nervous system (126). However, our study was unable to identify any transcriptional response to IL-23. This might be explained by power limitations, especially given our small sample size and the fact that the biological replicates of IL-23 stimulation appear quite separate from each other in the PCA plot. However, it is also possible that this cytokine might act in a different cell type, for example in the polarisation of Th17 cells (25, 203).

Even though macrophages are not considered drivers of autoimmunity, their role in several immune diseases is well established (4, 87, 204). As a result, we analysed the gene expression profile of macrophages polarised to M1 and M2 phenotypes. We observed transcriptional responses with 460 differentially expressed genes only observed in M1, 89 only in M2, and 25 genes shared. A comparison of their enrichments in GO terms revealed that the M1 response might be coordinated by STAT1 and IRFs. On the other hand, STAT6 could be coordinating the M2 response, which agrees with the literature (87, 92). This duality of transcriptional regulation is mirrored at the functional level, with M1 macrophages showing upregulation of chemokine receptors and genes necessary for the antiviral response (RNAses and OAS), and M2 macrophages downregulating genes in the IL-1 and IL-6 synthesis and signalling pathways. These observations agree with the generally accepted idea of M2 macrophages being involved in tissue repair and wound healing (83, 87).

In summary, there appear to be two patterns of transcriptional dynamics upon cytokine induced polarisation of $CD4^+$ T cells: steadily increasing changes in gene expression (known lineages) and transitional changes, (IFN-β). We identified pathways which are relevant for

the function of each of these cell states using pathway enrichment analysis followed by overrepresentation analysis. For instance, we observed that downregulation of the PI3K pathway might be involved in Th17 and iTreg differentiation and that IFN-β might induce components of the antiviral response, among other examples. Furthermore, we hypothesised possible mechanisms of gene regulation using gene co-expression network analysis such as downregulation of IRFs in Th2 cells and Sox4 in Th17 cells. Last, we also characterised the transcriptome of known macrophage polarisation states and proposed that the effect of IL-26 could act via the TF NFκB, sharing some aspects with TNF-α polarisation.