

# **A NOVEL ROLE FOR ZEB2 AS A LINEAGE FIDELITY CHECKPOINT IN HUMAN CD4+ T CELLS**

SOON WEI WONG

Discipline of Paediatrics, School of Medicine,  
University of Adelaide

MAY 2021

A thesis submitted to the University of Adelaide as the requirement  
for the degree of Doctor of Philosophy

## Table of Contents

Table of Contents.....	i
List of Tables.....	x
List of Figures.....	xiv
Abstract.....	xx
Declaration.....	xxii
Acknowledgements .....	xxiv
Abbreviations.....	xxvi
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
1.1 Immune System .....	2
1.2 Cells of The Immune System.....	5
1.2.1 Antigen Presenting cells.....	5
1.2.2 T cells and B cells .....	6
1.3 Immune Homeostasis and Tolerance .....	8
1.4 CD4+ T cell Lineages Differentiation .....	9
1.5 Mechanisms of Treg Action.....	12
1.6 Thymic Treg and Peripheral Treg.....	14
1.7 Mirrored Suppressive and Effector Responses .....	17
1.8 Molecular Mechanisms of FOXP3 Function .....	20
1.9 ZEB2 – A Key Gene Repressed by FOXP3 in Treg cells .....	21
1.10 Molecular Structure and Transcriptional Activities of ZEB2.....	24
1.11 Clinical Significance of ZEB2 .....	25
1.12 ZEB2 Regulates Functional Maturation in Immune Cells.....	26

1.13	ZEB2 is expressed in Th1 Cells.....	29
1.14	The Unanswered question: Why is ZEB2 in Expressed in human CD4+ T Effector Cells?.....	32
1.15	Hypothesis and Aims .....	34
<b>CHAPTER 2: MATERIALS &amp; METHODS .....</b>		<b>35</b>
2.1	T Cell Isolation, Culture and Antibody Labelling of Proteins.....	36
2.1.1	Isolation of peripheral blood mononuclear cells .....	36
2.1.2	CD4+ T cells isolation.....	37
2.1.3	Enrichment of Treg/Tconv or Memory/Naïve populations using magnetic beads .....	39
2.1.4	Fluorochrome-conjugated antibody labelling of cell surface proteins.....	40
2.1.5	Fluorochrome-conjugated antibody labelling of nuclear proteins for flow cytometry.....	42
2.1.6	Sort purification and analysis of T cell populations by Flow cytometry .....	43
2.1.7	CD4+ T cell culture.....	44
2.1.8	Th1 polarisation.....	45
2.1.9	Analysis of intracellular cytokines .....	45
2.2	Molecular Biology .....	46
2.2.1	Isolation of RNA .....	46
2.2.2	Nucleic acid quantification.....	47
2.2.3	cDNA synthesis.....	48
2.2.4	Quantitative Real-Time PCR Reactions and Analysis.....	48

2.3	CRISPR/Cas9 Knockout.....	51
2.3.1	CRISPR Design.....	51
2.3.2	Nucleofection .....	52
2.3.3	Single Cell Clone Expansion.....	53
2.3.4	Determination of Knockout Efficiency .....	54
2.3.5	Sequencing of ZEB2 Genomic DNA Targeted by sgRNAs .....	56
2.3.6	Inference of CRISPR Edits (ICE) Analysis .....	56
2.4	RNA-Sequencing .....	58
2.4.1	Assessment of RNA Quality Assessment for RNA Sequencing.....	58
2.4.2	Selection of Poly(A) mRNA .....	59
2.4.3	RNA-seq library generation .....	59
2.4.4	Quality Control Assessment.....	62
2.4.5	Pseudoalignment .....	62
2.4.6	Differential Gene Expression Analysis .....	62
2.5	Overexpression of ZEB2 with lentivirus .....	63
2.5.1	Plasmids for packaging HEK293T/17 cells with LV411 lentivirus.....	63
2.5.2	Plasmid Purification .....	64
2.5.3	Thawing of HEK293T/17 Cells .....	65
2.5.4	HEK293T/17 Cell Culture and Preparation .....	65
2.5.5	Packaging and Concentration of Lentiviral Supernatant.....	66
2.5.6	Viral Titre .....	68
2.5.7	Transduction of CD4+ T Cells with LV411-ZEB2-IRES-hrGFP.....	69

<b>CHAPTER 3: ZEB2 EXPRESSION IN CD4+ T CELLS.....</b>	<b>70</b>
3.1 Introduction.....	71
3.2 Aims and Hypothesis .....	74
3.3 Material & Methods.....	75
3.3.1 ZEB2 Gene Locus Metagenomics.....	75
3.3.2 Detection of ZEB2 by flow cytometry in ZEB2 overexpressing HEK293T/17 cells.....	75
3.3.3 Flow Cytometry Panels for Isolating CD4+ T Cell Subsets .....	76
3.3.4 Detection of ZEB2 by Flow Cytometry in Treg and Tconv .....	80
3.3.5 UMAP Dimensionality Reduction of T Helper Lineage Populations.....	81
3.3.6 Central and Effector Memory Helper Lineage Activation.....	82
3.4 Results.....	83
3.4.1 ZEB2 Gene Locus Accessibility .....	83
3.4.2 Detection of ZEB2 Protein Expression by Flow Cytometry .....	84
<i>ZEB2 antibody labelling in HEK293T/17 cell lines overexpressing ZEB2.....</i>	<i>84</i>
<i>ZEB2 antibody labelling in Treg and Tconv .....</i>	<i>87</i>
3.4.3 ZEB2 is Highly Expressed in Effector Memory CD4+ T Cells.....	91
3.4.4 ZEB2 Expression in CD25 MACS Enriched Purified Treg and Tconv Helper Lineages .....	97
3.4.5 ZEB2 Expression in CD45RA MACS Enriched Purified Treg and Tconv Helper Lineages .....	106
<i>Comparison of the helper lineages proportions in Treg and Tconv.....</i>	<i>106</i>

	<i>ZEB2 mRNA expression in Treg and Tconv helper lineage populations</i> .....	107
3.4.6	ZEB2 Expression in CM and EM of Purified Tconv Helper Lineage Populations .....	120
	<i>Characterising Central and Effector memory of Tconv helper lineage populations</i> .....	120
	<i>Dimension reduction of Memory Tconv pools with UMAP</i> .....	123
	<i>ZEB2 expression across memory subsets of Tconv helper lineage populations</i> .....	127
3.4.7	Characterising ZEB2 expression in CM and EM following activation .....	134
3.5	Discussion .....	144
<b>CHAPTER 4: ZEB2 FUNCTION IN TH1 EFFECTOR MEMORY CELLS.....</b>		<b>150</b>
4.1	Introduction.....	151
4.2	Aims and Hypothesis .....	154
4.3	Material & Methods .....	155
4.3.1	Cell Isolation .....	155
4.3.2	CRISPR/Cas9 Deletion of ZEB2 .....	156
	<i>Preparation of cells</i> .....	156
	<i>Pre-complexing sgRNA-Cas9 to form Ribonucleoprotein (RNP)</i> .....	156
	<i>Nucleofection</i> .....	157
	<i>Determination of knockout efficiency</i> .....	158
4.3.3	RNA-seq.....	159
	<i>Library preparation</i> .....	159

	<i>Statistical analysis</i> .....	159
4.3.4	Lentivirus overexpression of ZEB2 .....	159
	<i>Preparation of cells</i> .....	159
	<i>Lentivirus transduction of T cells</i> .....	160
4.3.5	Quantitative Real Time PCR.....	160
4.3.6	Th1 polarisation.....	161
4.4	Results.....	162
4.4.1	Optimising CRISPR/Cas9 techniques to knockout ZEB2 in Tconv Th1 EM	162
4.4.2	A single sgRNA CRISPR/Cas9 deletion of ZEB2.....	162
	<i>Knockout efficiency verification</i> .....	162
	<i>Genotype and expression profiles of ZEB2 deleted clones</i> .....	165
	<i>Comparison of gene expression in Th1 EM ZEB2 knockout and wild-type clones by RNA-seq</i> .....	168
4.4.3	A Triple sgRNAs CRISPR/Cas9 Deletion of ZEB2 .....	175
	<i>Analysis of ZEB2 knockout in pooled Th1 EM cells using 3 sgRNAs strategy</i> .....	175
	<i>ZEB2 is required for the expression of Th1 effector cytokine, IFN<math>\gamma</math></i> .....	178
	<i>Th1 lineage fidelity is maintained by ZEB2</i> .....	181
4.4.4	ZEB2 is not required for formation of Th1 .....	188
4.4.5	Common ZEB2 DE between Clones and Pool.....	190
4.4.6	ZEB2 overexpression in Naïve Tconv and Treg .....	192
4.5	Discussion .....	195

<b>CHAPTER 5: META-ANALYSIS OF ZEB2 RNA-SEQ DATA .....</b>	<b>200</b>
5.1 Introduction.....	201
5.2 Aims and Hypothesis .....	204
5.3 Material & Methods.....	205
5.3.1 Gene Ontology (GO) Functional Annotation Analyses .....	205
5.3.2 KEGG Pathway Functional Annotation Analyses .....	205
5.3.3 Gene Set Enrichment Analysis.....	205
5.3.4 Enrichr GWAS 2019 Analysis .....	206
5.3.5 ZEB2 ChIP Filtered by ATACseq .....	206
<i>Global ZEB2 ChIP-seq.....</i>	<i>206</i>
<i>Th1 ATAC-seq.....</i>	<i>207</i>
<i>Annotation and Intersection .....</i>	<i>207</i>
5.3.6 Gene Co-expression Network Analysis with CEMiTool.....	208
5.4 Results.....	210
5.4.1 Enrichment Analysis for ZEB2 KO Th1 EM Clones.....	210
<i>Gene Ontology annotation.....</i>	<i>210</i>
<i>KEGG pathway annotation .....</i>	<i>216</i>
5.4.2 Enrichment Analysis for ZEB2 KO Th1 EM Pools.....	218
<i>Gene Ontology annotation with DE list of FDR &lt; 0.05.....</i>	<i>218</i>
<i>KEGG pathway annotation with DE list of FDR &lt; 0.05.....</i>	<i>223</i>
<i>Gene Ontology annotation with DE list of FDR &lt; 0.1.....</i>	<i>225</i>
<i>KEGG pathway annotation with DE list of FDR &lt; 0.1.....</i>	<i>232</i>



<i>Gene set enrichment analysis with DE list of FDR &lt; 0.1</i> .....	235
5.4.3 ZEB2 ChIP gene target in Th1 .....	237
5.4.4 ZEB2 regulates gene networks to promote Th1 effector memory cell transcriptional programming. ....	244
5.4.5 Gene co-expression network analysis with CEMiTool.....	248
5.4.6 Enrichr: GWAS annotation with DE list of FDR < 0.1 .....	259
5.5 Discussion.....	262
<b>CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS.....</b>	<b>268</b>
6.1 Discussion.....	269
6.2 Future direction.....	277
6.3 Conclusions.....	280
<b>CHAPTER 7: APPENDIX.....</b>	<b>282</b>
7.1 RNA-seq Libraries Preparation Summaries.....	283
7.2 Next Generation Sequencing (NGS) Reports .....	284
7.2.1 ngsReports: ZEB2 Knockout in Th1 CM & EM Clones .....	284
7.2.2 ngsReports: ZEB2 Knockout in Th1 CM & EM Pools.....	292
7.3 RNA-seq bash Scripts .....	306
7.3.1 Bash script for ZEB2 knockout in Th1 EM clones .....	306
<i>Pipeline script</i> .....	306
<i>Additional script for Salmon</i> .....	310
7.3.2 Bash script for ZEB2 knockout in Th1 EM pools.....	311
<i>Pipeline script</i> .....	311

<i>Additional script for Salmon</i> .....	314
7.3.3 Bash script for activated Th1 ATAC-seq.....	315
7.4 RNA-seq R Scripts.....	317
7.4.1 Differential Expression Analysis .....	317
7.4.2 Enrichment Analysis .....	328
<i>Gene Ontology and KEGG pathway Analysis</i> .....	328
<i>Gene set enrichment analysis</i> .....	335
<b>REFERENCES</b> .....	<b>340</b>

## List of Tables

Table 1.1: Comparison between innate and adaptive immunity .....	4
Table 2.1: Antibodies used for labelling surface protein for Flow Cytometry .....	40
Table 2.2: Antibodies used for labelling nuclear protein for Flow Cytometry .....	42
Table 2.3: Primers used in qRT-PCR experiments .....	50
Table 2.4: sgRNAs sequence used in CRISPR/Cas9 ZEB2 knockout experiments .....	52
Table 2.5: Thermal cycler program for genomic DNA extraction.....	54
Table 2.6: PCR primers sequence used in CRISPR/Cas9 ZEB2 knockout experiments .....	55
Table 2.7: Preparation guide for PCR reaction .....	55
Table 2.8: PCR thermocycler program .....	55
Table 2.9: Recommended amounts of DNA and primer for sequencing reactions.....	56
Table 3.1: Panel 1: Treg and Tconv panel .....	77
Table 3.2: Panel 2: Naïve & memory panel .....	77
Table 3.3: Panel 3 – Chemokine receptors panel .....	78
Table 3.4: Panel 4 – Memory chemokine receptor panel.....	79
Table 3.5: Experimental FACS tubes for ZEB2 antibody labelling in Treg and Tconv .....	81
Table 3.6: Anova Fisher’s LSD test on ZEB2 expression in memory subset profile of Treg and Tconv.....	94
Table 3.7: Anova Fisher’s LSD test on T-bet expression in memory subset profile of Treg and Tconv.....	95
Table 3.8: Anova Fisher’s LSD test on FOXP3 expression in memory subset profile of Treg and Tconv .....	96
Table 3.9: Anova Fisher’s LSD test on ZEB2 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment .....	102

Table 3.10: Anova Fisher’s LSD test on FOXP3 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment .....	103
Table 3.11: Anova Fisher’s LSD test on T-bet expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment .....	104
Table 3.12: Anova Fisher’s LSD test on GATA3 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment .....	105
Table 3.13: Anova Fisher’s LSD test on ZEB2 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment .....	110
Table 3.14: Anova Fisher’s LSD test on T-bet expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment .....	112
Table 3.15: Anova Fisher’s LSD test on FOXP3 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment.....	114
Table 3.16: Anova Fisher’s LSD test on ROR $\gamma$ t expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment.....	116
Table 3.17: Anova Fisher’s LSD test on GATA3 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment.....	118
Table 3.18: Mean and SEM of Effector Memory and Central Memory subsets proportion in each Tconv and Treg helper lineage populations.....	122
Table 3.19: Proportion of UMAP clusters with population identity classified by chemokine receptor expression.....	127
Table 3.20: Anova Fisher’s LSD test on ZEB2 expression in memory subsets of Tconv helper lineage populations.....	130
Table 3.21: Anova Fisher’s LSD test on T-bet expression in memory subsets of Tconv helper lineage populations.....	131

Table 3.22: Anova Fisher’s LSD test on ROR $\gamma$ t expression in memory subsets of Tconv helper lineage populations.....	132
Table 3.23: Anova Fisher’s LSD test on GATA3 expression in memory subsets of Tconv helper lineage populations.....	133
Table 3.24: Anova Fisher’s LSD test on ZEB2 expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells.....	136
Table 3.25: Anova Fisher’s LSD test on T-bet expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells.....	138
Table 3.26: Anova Fisher’s LSD test on GATA3 expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells.....	140
Table 3.27: Anova Fisher’s LSD test on ROR $\gamma$ t expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells.....	142
Table 4.1: Top 35 upregulated genes in ZEB2 KO Th1 EM clones.....	171
Table 4.2: Top 35 downregulated genes in ZEB2 KO Th1 EM clones.....	172
Table 4.3: Summary of ICE analysis of triple sgRNAs ZEB2 KO Th1 EM.....	178
Table 4.4: Top 35 upregulated genes in ZEB2 KO Th1 EM pools.....	185
Table 4.5: Top 35 downregulated genes in ZEB2 KO Th1 EM pools.....	186
Table 5.1: Significantly enriched GO terms using an FDR < 0.05 and using ZEB2 KO Th1 EM clones DE genes with FDR < 0.05.....	213
Table 5.2: Significantly enriched KEGG terms using a p-value < 0.05 and using ZEB2 KO Th1 EM clones DE genes with FDR < 0.05.....	216

Table 5.3: Significantly enriched GO terms using an FDR < 0.1 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.05 .....	220
Table 5.4 Significantly enriched KEGG terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.05 .....	223
Table 5.5: Significantly enriched GO terms using an FDR < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1 .....	227
Table 5.6: Significantly enriched KEGG terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1 .....	233
Table 5.7: Direct targets of ZEB2 transcription factor .....	239
Table 5.8: Modules enrichment score with adjusted p-value .....	251
Table 5.9: Modules information report .....	251
Table 5.10: Genes in M1 module .....	251
Table 5.11: Genes in M2 modules .....	252
Table 5.12: Genes in M3 modules .....	252
Table 5.13: Significantly enriched GWAS terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1 .....	260
Table 7.1: RNA-seq libraries preparation summary of ZEB2 knockout in Th1 EM clones from donors #070618 & #150618 .....	283
Table 7.2: RNA-seq libraries preparation summary of ZEB2 knockout in Th1 CM & EM pools from donors #08, #10, #16 & #18. ....	283

## List of Figures

Figure 1.1: Immune cells involve in the Innate and adaptive immunity .....	6
Figure 1.2: Antigen presentation stimulates CD8+ cytotoxic cells or CD4+ helper T cells to mature .....	7
Figure 1.3: Schematic illustrates a fine balance between tolerance and reactivity of healthy immune system.....	9
Figure 1.4: CD4+ T cell subset differentiation .....	11
Figure 1.5: Mechanisms of Treg suppression .....	13
Figure 1.6: Phenotypically mirrored Treg and Tconv helper lineages.....	19
Figure 1.7: ZEB2 is a bona fide target of FOXP3 and miR-155.....	22
Figure 1.8: Expression and regulation of ZEB2 in Treg and Tconv by FOXP3 and miR-155	23
Figure 1.9 ZEB2 could act as either transcriptional repressors or activators depending on the target gene and tissue .....	24
Figure 1.10: Summary of ZEB2 roles in DC, macrophages, B-cell, CD8+ T cell, NK cell and CD4+ T cell.....	28
Figure 1.11: ZEB2 and T-bet expression is high in the effector memory of CD4+ T cell .....	29
Figure 1.12: CD4+ T cell memory phenotype panel.....	30
Figure 1.13: ZEB2 and T-bet expression is higher in Th1 compared with Th17 in mouse.....	31
Figure 2.1: Purity of enriched CD4 T cells .....	38
Figure 2.2: Lymphocyte gating for flow cytometry .....	43
Figure 2.3: Schematic of 3 sgRNAs and PCR screening primers around ZEB2 exon 3 .....	52
Figure 2.4: ICE summary window .....	57
Figure 2.5: Electropherogram of an RNA sample used for RNA-seq .....	59
Figure 2.6: RNA-seq library size distribution on an Experion™ Automated Electrophoresis System .....	61

Figure 2.7: Components of third-generation lentiviral plasmids .....	64
Figure 2.8: LV411-ZEB2-IRES-hrGFP transfer plasmid construct.....	64
Figure 2.9: LV411-eGFP-IRES-hrGFP transfer plasmid construct .....	64
Figure 2.10: Proportion of GFP+ HEK 293T cells transduced with lentivirus.....	69
Figure 3.1: Visualizing of ZEB2 gene locus of Treg and Tconv .....	83
Figure 3.2: ZEB2 mRNA expression in HEK293T & HEK293T overexpressing ZEB2.....	85
Figure 3.3: Analysis of ZEB2 protein level in HEK293T & HEK293T overexpressing ZEB2 .....	86
Figure 3.4: FOXP3 and ZEB2 mRNA expression in Treg and Tconv .....	88
Figure 3.5: Physical gating strategy for Treg and Tconv .....	89
Figure 3.6: ZEB2 protein level in Treg and Tconv .....	90
Figure 3.7: Memory subset profiling of Treg and Tconv.....	92
Figure 3.8: ZEB2, T-bet and FOXP3 mRNA expression in memory subset profile of Treg and Tconv.....	93
Figure 3.9: CD25 MACS purity FACS plot.....	98
Figure 3.10: Chemokine receptor helper lineages profiling and sorting strategy of post CD25 MACS Treg and Tconv .....	99
Figure 3.11: ZEB2, T-bet and FOXP3 mRNA expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment.....	101
Figure 3.12: Chemokine receptor helper lineages profiling and sorting strategy from post CD45RA- enriched MACS CD4+ T cell .....	106
Figure 3.13: Proportion of Tconv and Treg helper lineage populations .....	107
Figure 3.14: ZEB2, T-bet, ROR $\gamma$ t and FOXP3 mRNA expression in helper lineage populations of Treg and Tconv .....	109



Figure 3.15: Gating strategy of Central Memory and Effector memory from Treg and Tconv helper lineage populations.....	121
Figure 3.16: Proportion of Effector Memory and Central Memory subsets in Tconv and Treg helper lineage populations.....	122
Figure 3.17: UMAP of memory Tconv stained with the chemokine receptor panel .....	125
Figure 3.18: MFI Heatmap of UMAP clusters.....	126
Figure 3.19: ZEB2, T-bet, GATA3 and ROR $\gamma$ t mRNA expression in memory subsets of Tconv helper lineage population .....	129
Figure 3.20: ZEB2 (blue), T-bet (green), GATA3 (orange) and ROR $\gamma$ t (red) mRNA expression in non-activated (unstim) or activated (stim) Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 .....	135
Figure 4.1: Schematic of single CRISPR guide targeting ZEB2 exon 3.....	163
Figure 4.2: Gel image of genomic PCR post CRISPR/Cas9 using single sgRNA .....	164
Figure 4.3: ICE analysis of single sgRNAs ZEB2 KO Th1 EM prior to single cell cloning.	165
Figure 4.4: Genotype of ZEB2 KO clones by ICE analysis.....	167
Figure 4.5: Gene expression changes of ZEB2 KO Th1 EM clones by qRT-PCR.....	168
Figure 4.6: PCA of ZEB2 KO & WT Th1 EM clones before and after RUVg normalisation .....	169
Figure 4.7: Volcano plot comparing ZEB2 KO vs WT Th1 EM clones.....	170
Figure 4.8: Heatmap of top 100 DE genes between ZEB2 KO and WT Th1 EM clones.....	173
Figure 4.9: Schematic of 3 sgRNAs targeting exon 3 of ZEB2 gene .....	175
Figure 4.10: Segregation of post-nucleofected Th1 into CM and EM.....	177
Figure 4.11: Gel image of genomic PCR post CRISPR/Cas9 using three sgRNAs .....	177
Figure 4.12: Gene expression changes in ZEB2 KO Th1 CM and EM cells by qRT-PCR...	180

Figure 4.13: PCA of ZEB2 KO & WT Th1 EM cell pools before and after RUVg normalisation .....	183
Figure 4.14: Volcano plot comparing ZEB2 KO vs WT Th1 EM pools .....	184
Figure 4.15: Heatmap of top 100 DE genes between ZEB2 KO and WT Th1 pools .....	187
Figure 4.16: Gel image of genomic PCR post ZEB2 deletion in Naïve Tconv .....	189
Figure 4.17: ZEB2 KO Naïve Tconv polarised to Th1 .....	189
Figure 4.18: Overlapping differential expressed (DE) genes from ZEB2 KO clones and ZEB2 KO pool of Th1 EM cells .....	190
Figure 4.19: Fold change difference of common genes between ZEB2 KO clones and ZEB2 KO pool of Th1 EM cells dataset .....	191
Figure 4.20: Gene expression changes of ZEB2 overexpression in Naïve Tconv and Treg..	194
Figure 5.1: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM clones .....	212
Figure 5.2: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM clones.....	214
Figure 5.3: Gene-concept network plot showing all Molecular Function GO terms from Th1 EM clones.....	215
Figure 5.4: Gene-concept network plot showing top 12 significantly enriched KEGG pathway terms from Th1 EM clones.....	217
Figure 5.5: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM pools (Gene list FDR < 0.05).....	219
Figure 5.6: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM pools (Gene list FDR < 0.05) .....	221
Figure 5.7: Gene-concept network plot showing all Molecular Function GO terms .....	222

Figure 5.8: Gene-concept network plot showing all significantly enriched KEGG pathway terms from Th1 EM pools (Gene list FDR < 0.05) .....	224
Figure 5.9: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM pools (Gene list FDR < 0.1) .....	226
Figure 5.10: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM pools (Gene list FDR < 0.1) .....	229
Figure 5.11: Gene-concept network plot showing all Molecular Function GO terms from Th1 EM pools (Gene list FDR < 0.1) .....	230
Figure 5.12: Gene-concept network plot showing all Cellular Component GO terms from Th1 EM pools (Gene list FDR < 0.1) .....	231
Figure 5.13: Gene-concept network plot showing top 20 significantly enriched KEGG pathway terms from Th1 EM pools (Gene list FDR < 0.1) .....	234
Figure 5.14: Barcode plot of enriched Hallmark gene sets with GSEA analysis.....	236
Figure 5.15: Venn diagram showing number of genes that are a target of ZEB2.....	238
Figure 5.16: Functional and molecular specialization of DE genes of ZEB2 deleted Th1 EM pools with annotated ZEB2 ChIP .....	243
Figure 5.17: PCA of Th1 CM & Th1 EM cells.....	245
Figure 5.18: Volcano plot comparing Th1 CM vs Th1 EM cells .....	246
Figure 5.19: ZEB2 regulates a subset of EM and CM signature genes .....	247
Figure 5.20: Heatmap of CM and EM signature genes that are regulated by ZEB2 .....	247
Figure 5.21: Sample clustering tree (dendrogram).....	249
Figure 5.22: Beta x R <sup>2</sup> plot and Mean connectivity plot. ....	249
Figure 5.23: Heatmap of module enrichment analysis.....	253
Figure 5.24: Over Representation Analysis of modules M1 using gene sets from the MsigDB Hallmark Pathway database. ....	254

Figure 5.25: Over Representation Analysis of modules M2 using gene sets from the MsigDB Hallmark Pathway database. ....	255
Figure 5.26: Over Representation Analysis of modules M3 using gene sets from the MsigDB Hallmark Pathway database. ....	256
Figure 5.27: Interaction plot for M1, with gene nodes highlighted. ....	257
Figure 5.28: Interaction plot for M3, with gene nodes highlighted. ....	258
Figure 5.29: Gene-concept network plot of IBD related autoimmune diseases from Enrichr analysis using Th1 EM pools (Gene list FDR < 0.1) .....	261
Figure 6.1: ZEB2 role and regulation in Treg paradigm.....	279
Figure 6.2: ZEB2 regulation and function in CD4+ T cell in implicating IBD pathogenesis	281

## Abstract

Autoimmune diseases are a broad range of more than eighty related disorders, affecting up to 5% of the population. The incidence of autoimmune disease is increasing worldwide. It is a disease where the body's immune system fails to recognize its own cells and tissues as “self”. Instead, immune cells attack these healthy cells and tissues as if they were foreign or invading pathogens. One of the key immune cell populations implicated in this immune attack is CD4<sup>+</sup> T cells. The CD4<sup>+</sup> T cell lineage consists of a number of phenotypically and functionally distinct subsets. In particular there are two functionally distinct compartments in CD4, namely T regulatory cells (Treg) and T conventional cells (Tconv), and the function of each is potentially altered in autoimmune disease. My PhD project has investigated the role of a transcription factor, ZEB2 in shaping the function of human CD4<sup>+</sup> T cells. Little is known about the role of ZEB2 in CD4<sup>+</sup> T cells and therefore elucidating its role in CD4<sup>+</sup> T cells and identifying the transcriptional landscape controlled by ZEB2 has the potential to highlight novel targets for autoimmune disease diagnosis and therapy.

ZEB2 is a zinc-finger transcription factor known to play a major role in early embryogenesis and in tumour metastasis. ZEB2 has an established role in the cancer metastasis of several cancers but its role in the immune system has only fairly recently been explored. Interestingly, ZEB2, is directly induced by T-bet (T helper 1 master transcription factor) in mouse NK cells and CD8<sup>+</sup> T cells, and therefore I speculated that T-bet may be implicated in the regulation of ZEB2 in CD4<sup>+</sup> T cells where T-bet is the defining transcription factor for Th1 cells.

My PhD project identifies which CD4<sup>+</sup> T cell subsets ZEB2 is expressed in. I show that ZEB2 is expressed highly in Tconv effector memory subsets, indicating its role in the effector compartment of CD4<sup>+</sup> T cells. Further investigation indicated that ZEB2 was found predominantly in Th1 effector memory (EM) cells. ZEB2 was expressed at very low levels in the other Tconv helper lineages, suggesting a unique effector role of ZEB2 in Th1 where T-bet

is highly expressed and FOXP3 is absent. However, the regulation of ZEB2 is clearly more complex, since in some CD4<sup>+</sup> T cell subsets with high T-bet, for instance Th1/17, there is not necessarily high ZEB2, suggesting ZEB2 is not regulated by T-bet alone.

In order to specifically define the role of ZEB2 in Th1 EM cells, I deleted ZEB2 and analysed global changes in gene expression by RNA-seq. RNA-seq analysis showed that 222 genes were differentially expressed between WT and ZEB2-deleted Th1 EM, and pathway analysis of the gene profile indicates a potential role for ZEB2 in regulating inflammatory cytokines, repressing cytotoxic responses, enhancing motility and increasing survival in high stress environments. ZEB2 is also shown to regulate effector memory and central memory genes important for Th1 effector memory differentiation. Hence, ZEB2 is important in maintaining the function and fidelity of a Th1 effector memory cell in the steady state, and indirectly or directly maintaining IFN $\gamma$  expression. Th1 cells preferentially produce IFN $\gamma$  and IL-2 and are the principal regulators of type 1 immunity (Th1 response), which eradicates intracellular pathogens including viruses. Unravelling the role of ZEB2 in the complex relationships between the Th1 and Treg lineages and subsets may provide critical insight into the disruption of immune homeostasis that leads to autoimmune disease including inflammatory bowel disease (IBD), and may suggest novel therapeutic targets for autoimmune diseases.

## Declaration

I, Soon Wei WONG, author of this thesis entitled, “A Novel Role for ZEB2 as a Lineage Fidelity Checkpoint in Human CD4+ T Cells”, certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of a Beacon of Enlightenment PhD Scholarship.

Soon Wei Wong

May 2021

“It is good to love many things, for therein lies the true strength, and whosoever loves much performs much, and can accomplish much, and what is done in love is well done.”

Vincent Van Gogh



## Acknowledgements

First and foremost, I am extremely grateful to my supervisor, Professor Simon Barry for supporting me during these past 5 years since honours. Simon is someone you will instantly love and never forget once you meet him. He always has been non-judgmental of me and instrumental in instilling confidence. He has faith in me and my intellect even when I didn't have faith in myself. Thank you for agreeing to have me as your student and for putting me on the path to a promising future! I would also like to thank my co-supervisor Dr. Cheryl Brown for her support and teaching me the necessary laboratory skills for my research, as well as always being willing to answer my science questions at all times. She has been supportive and has given me the freedom to pursue my own experimental approach in my PhD projects. She has also provided insightful discussions about the research findings from my data which I am very grateful of. I am also very thankful to my co-supervisor Dr. Timothy Sadlon for his scientific advice and knowledge including many insightful discussions and suggestions on my CRISPR experiments and my RNA-seq data analysis.

I would also like to thank all the present members of the Molecular Immunology group: Ying who has been a great company in bioinformatics, so I know I am not alone here in this lab; Chris who has been very helpful with flow cytometry and panel design during the start of my PhD and with all the countless sorts and staining; Veronika who has always been nice, helpful and supportive during my search for jobs and thank you for all the good reference letters you wrote about me; Baggy who has always been a good neighbour sitting next to me and helped me with understanding every aspect of flow cytometry; Katherine who I teased a lot and I meant no harm to that. She is very nice and helpful, and I hope that all the things that I have taught will be useful for you in the future; and thanks to other members of the lab: Jacquie, Holly, Kate, Leila and Emma. I also want to thank the past members of group; Silvia who helped me with ordering reagents and Kristen who has been very helpful when I was an honours student.

I also want to thank Dr. Claudia Bossen for hosting me while I was in Freiburg, Germany and her PhD student Neftali Ramirez, who has been very kind and friendly, and showed me the inside world of bioinformatics. Even though the University of Freiburg joint program did not work out in the end, I am very thankful for being given the opportunity, as I would not be doing bioinformatics without this trip. I also want to thank Dr. Steve Pederson for assisting me with all my bioinformatic questions and providing some of the scripts I have used in my PhD.

I would also like to thank House of Chow restaurant for providing free meals and additional income on top of my stipend while I was working there. I am grateful to the owner of the restaurant, Roz who was very flexible with my availability to work and made sure I was well fed before I started my shift. To the manager, Paul who has always looked out for me while I'm working. I really enjoyed being part of the team, and it is also quite fun to work with people outside of the research field.

The best outcome from these past five years is finding my partner, Karen. Thank you for believing in me, for encouraging me up till the end of my PhD. I don't think I could have made it through without your company all along this journey. You chose to stay with me in Adelaide even though that wasn't your plan at all. You also sacrificed a better career opportunity in Hong Kong for me. There are no words to convey how much I love you. Now this is over, I promise to finally get a job and make time to take the next step in our relationship ♡. To my family back in Malaysia who are eagerly waiting for me to complete my PhD, here you go, I am finally finished! My hard-working parents have sacrificed their lives for me so I can study abroad and provided unconditional love and care. I love them so much, and I would not have made it this far in life without them. I know I always have my family to count on when times are rough.

## Abbreviations

°C	degrees Celsius
µg	microgram
µL	microLitre
µM	microMolar
3'UTR	3 prime untranslated region
APC	Antigen Presenting Cells
ATAC-seq	Assay for Transposase-Accessible Chromatin sequencing
bp	base pairs
BP	Biological Process
Cat	Catalogue
CC	Cellular Component
CD	Crohn's Disease
cDNA	complimentary DNA
CEMiTool	Co-Expression Modules identification Tool
ChIP-seq	Chromatin Immunoprecipitation sequencing
CM	Central Memory
CO <sub>2</sub>	Carbon Dioxide
CPM	Counts Per Million
CRAC	Calcium Release Activated Chanel
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CTL	Cytotoxic T cells
DA	Differential Accesibile
DE	Differential Express
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EM	Effector Memory
EMRA+	Effector Memory RA+
EMT	Epithelial–mesenchymal transition
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FCS	Flow Cytometry Standard
FDR	False discovery rate

FOXP3	Forkhead box Protein3
FSC	Forward scatter
g	gravitational force
GFP	Green fluorescent protein
GLM	General Linear Model
GO	Gene Ontology
GWAS	Genome-wide association studies
GZM	granzyme
H <sub>2</sub> O	Water
HMOX1	Heme Oxygenase 1
hrGFP	Humanized Renilla reniformis green fluorescent protein
HS	High-Sensitivity
HSP	Heat shock proteins
IBD	Inflammatory bowel disease
ICE	Inference of CRISPR Edits
IDO	Indoleamine 2,3-dioxygenase
IFN	Interferons
Ig	Immunoglobulin
IL	Interleukin
INDEL	Insertion–deletion
IPEX	Immunodysregulation polyendocrinopathy enteropathy X-linked
IRES	Internal Ribosome Entry Site
iTreg	Induced T regulatory
kb	kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knockout
LSD	least significant difference
LV	Lentivirus
M	Molar
MACS	Magnetic-activated cell sorting
Mem	Memory
MF	Molecular Function
mg	miligram
MHC	Major histocompatibility complex

miR	microRNA
mL	miliLitre
MOI	Multiplicity of infection
mRNA	messenger Ribonucleuc acid
MS	Multiple Sclerosis
MSigDB	Molecular Signature Database
NF $\kappa$ B	Nuclear factor kappa B
ng	nanogram
NK	Natural Killer
nM	nanoMolar
ns	not significant
ORA	Over Representation Analysis
PAM	Protospacer adjacent motif
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PMA	Phorbol 12-Myristate 13-Acetate
pTreg	peripheral T regulatory
qRT-PCR	Quantitative Real Time PCR
RA	Rheumatoid arthritis
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RNP	Ribonucleoprotein
RT	room temperature
RUV	remove unwanted variation
scRNA-seq	single cell Ribonucleic acid sequencing
SEM	standard error of measurement
sgRNA	single guide RNA
SNP	Single Nucleotide Polymorphisms
SSC	Side scatter
T1D	Type 1 diabetes
T-bet	T-box protein expressed in T cells
Tconv	T conventional

TCR	T-cell receptor
TEMRA	terminally differentiated effector memory cells re-expressing CD45RA
TF	transcription factor
Tfh	T follicular helper
TGF $\beta$	Transforming growth factor beta
Th	T helper
TNF	Tumor necrosis factor
TNFSF	Tumor necrosis factor superfamily
TPM	Transcripts Per Million
Treg	T regulatory
TSDR	Treg-specific demethylated region
tTreg	thymic T regulatory
UC	Ulcerative Colitis
UMAP	Uniform Manifold Approximation and Projection
UV	ultraviolet
V	Voltage
WT	wildtype
ZEB2	Zinc Finger E-Box Binding Homeobox 2

**CHAPTER 1: INTRODUCTION**




## 1.1 Immune System

The human immune system is comprised of highly specialised cells that cooperate to defend the body from numerous pathogens and toxins in our environment. It is a complex and integrated system of cells, tissues, and organs that has specialized roles in defending against foreign antigens and pathogens, including bacteria, viruses, and fungi as well as guarding against cancer growth (Galli et al., 2020, Kellie and Al-Mansour, 2017). Together, the immune system provide both a first line of defence and an enduring recollection of threats, so that we are protected from re-exposure to pathogens we encounter (Marshall et al., 2018). In addition, this process is able to distinguish harmless challenges from harmful ones to prevent unnecessary immune activity, which can lead to disease (Gonzalez et al., 2011, Jiang and Chess, 2009). To achieve this, the immune system has developed a complex interacting network of cells and effector molecules that carry out constant surveillance of the immune landscape. When an immune challenge is first encountered, there is an immediate, albeit non-specific, immune response. This is carried out by tissues and cells which are loosely defined as belonging to the innate arm of the immune system and is essential, not only for initiating the removal of potential pathogens, but also for priming and activating cells that can mount a much more specific tailored immune response (the adaptive arm) (Marshall et al., 2018). Innate responses are elicited by a multitude of different mechanisms, including physical barriers (skin *etc.*) non-specific defence responses (mucous secretions *etc.*), platelet aggregation and activation of a number of different immune cells. There are several cell types that can provide this immediate and broad immune response characteristic of the innate system (neutrophils, monocytes, macrophages, dendritic cells and natural killer cells) and, additionally, some of these cells can present antigen to cells of the adaptive arm of the immune system (Cronkite and Strutt, 2018). Cells of the innate immune system cannot always recognize a pathogen or eliminate infectious organisms alone and also have no ability to elicit a more mature response against subsequent



reinfection, and so the immune system has evolved to have a more sophisticated and versatile system to respond to these challenges. This adaptive immune system comprises cells that provide the specificity and memory required to respond to the diverse immune challenges presented to the immune system and are crucial for immune memory responses. The adaptive immune system is organized into two classes of specialized lymphocytes, the T and B cells, which carry out several important and tightly linked functions (Cronkite and Strutt, 2018). They display an extremely diverse repertoire of antigen-specific recognition receptors that enable precise identification and elimination of pathogens, they carry out adaptive immune measures that ensure tailored immune responses and they provide long-lived immunological memory against reinfection. Thus, when a foreign pathogen is detected, an immune response is mounted by cells of the innate system which can then present antigen to, and activate cells of the adaptive immune system (summarised in Table 1.1 below). Once activated, the cells reprogram from a quiescent state to an activated state, initiating a cascade of signalling pathways resulting in changes in the expression of large numbers of genes and production of antigen specific cytokines, effector molecules, lipid mediators, tissue remodelling enzymes, reactive chemicals, and the ability to migrate through tissues and/or undergo clonal expansion (Marshall et al., 2018). The cells of the adaptive immune system can then retain an immunological memory to that antigen which will allow for a faster detection and more intense response against the same antigen in the future. Cells in the innate and adaptive compartment do not act independently, they interact with one another via chemical signalling molecules and together eliminate foreign pathogen (Lagou et al., 2018).

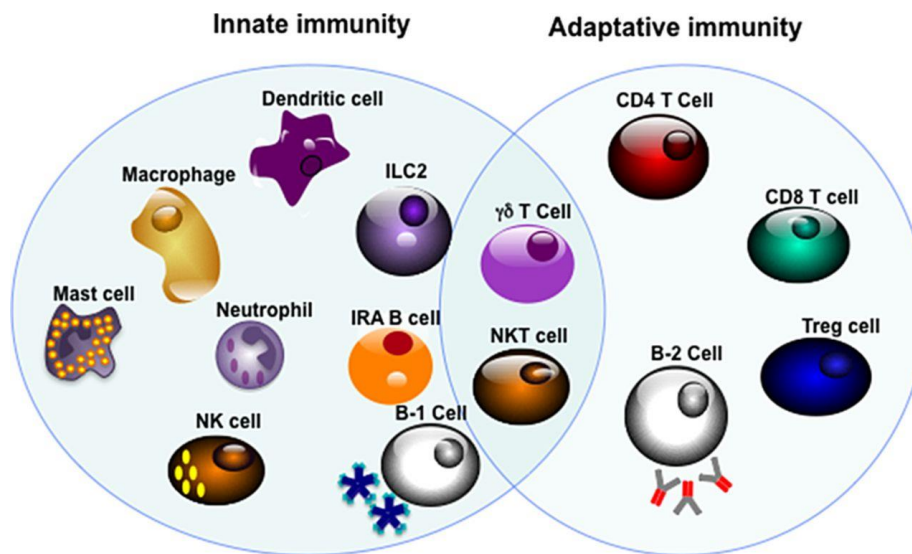
**Table 1.1: Comparison between innate and adaptive immunity**

	<b>Innate immunity</b>	<b>Adaptive Immunity</b>
<b>Components</b>	 <ol style="list-style-type: none"> <li>1. Physical and chemical barriers</li> <li>2. Phagocytic leukocytes</li> <li>3. Dendritic cells</li> <li>4. Natural Killer cells</li> <li>5. Plasma proteins (complement)</li> </ol>	 <ol style="list-style-type: none"> <li>1. Humoral immunity (B cells, which mature into antibody secreting plasma cells)</li> <li>2. Cell-mediated immunity (T cells, which mature into effector helper and cytotoxic T cells)</li> </ol>
<b>Activity</b>	Always present	Normally silent
<b>Response and potency</b>	Immediate response, but has a limited and lower potency	Slower response (over 1-2 weeks, but is much more potent)
<b>Specificity</b>	General: can recognize general classes of pathogens (i.e. bacteria, viruses, fungi, parasites) but cannot make fine distinctions	Recognizes highly specific antigens
<b>Course</b>	Attempts to immediately destroy the pathogen, and if it can't, it contains the infection until the more powerful adaptive immune system acts.	Slower to respond; effector cells are generally produced in 1 week and the entire response occurs over 1-2 weeks. However, this course can vary somewhat during different responses in an individual.
<b>Memory?</b>	No--reacts with equal potency upon repeated exposure to the same pathogen.	 Yes--memory cells "remember" specific pathogens; upon re-exposure to a pathogen, these cells mount a much faster and more potent second response

## 1.2 Cells of The Immune System

### 1.2.1 Antigen Presenting cells

The immune system consists of different immune cell types categorised into the innate and adaptive immune system as shown in Figure 1.1. The adaptive immune system must first identify each pathogen and then mount a response that will target that pathogen, without unwanted consequences to normal tissues and organs. The response is tailored to each type, eg: bacteria, viruses or parasite infection at different sites of the body and hence requires different mechanisms to eliminate different pathogens. One group of specialized cells that identify and present these foreign antigens for immune recognition are known as antigen presenting cells (APCs) (Eiz-Vesper and Schmetzer, 2020). The function of APCs is to display foreign antigens to lymphocytes. They have receptors that recognize a broad range of foreign antigens. Upon stimulation, the APC displays the antigen on its surface via major histocompatibility complex (MHC) class I or MHC class II molecules and simultaneously displays additional molecules on its surface that will act as co-activators of lymphocytes in the lymph node. MHC class I molecules are found on all nucleated cells; they present normal self-antigens as well as abnormal or non-self-pathogens to the effector T cells involved in cellular immunity. In contrast, MHC II molecules are only found on macrophages, dendritic cells, and B cells; they present abnormal or non-self-pathogen antigens for the initial activation of T cells. Dendritic cells are considered to be professional APCs (de Jong et al., 2006) but, other cells such as macrophages and B cells can also act as antigen presenting cells.

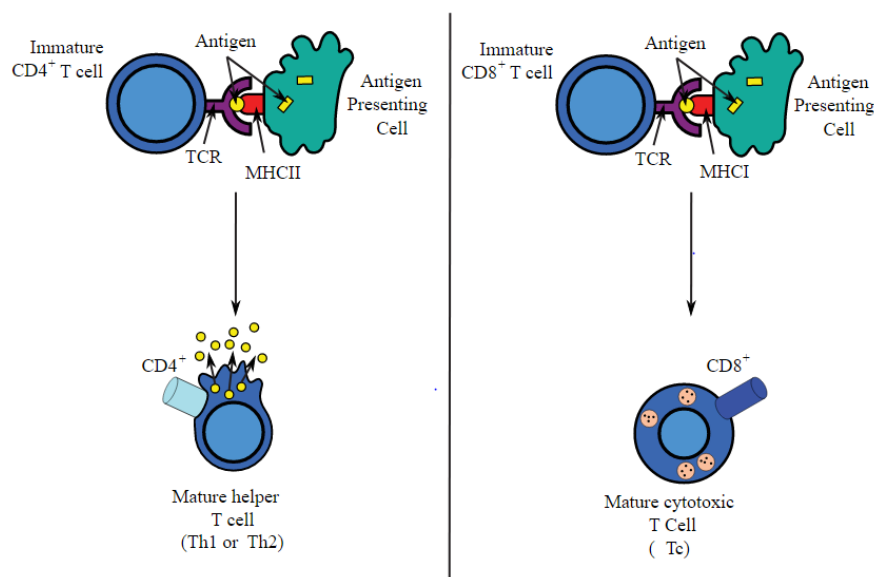


**Figure 1.1: Immune cells involve in the Innate and adaptive immunity**

### 1.2.2 T cells and B cells

There are two major cell types of the adaptive arm of the immune system. These cells are: B cells (also called B lymphocytes) and T cells (also called T lymphocytes). These lymphocytes have an array of highly specific receptors that recognize specific antigens in the body. B cells are derived from the bone marrow and generate antibodies when differentiated into plasma cells in the germinal centres. Highly specific antibodies are released to seek out and target antigens systemically, triggering the removal of the pathogens or infected cells. In contrast, T cells can directly kill infected host cells, activate other immune cells, produce cytokines and regulate the immune response. T cells mature in the thymus and express T cell receptors (TCR) that interact with MHC molecules on APCs during presentation of a specific antigen. There are two common types of T cells, the CD4<sup>+</sup> T helper cell and the CD8<sup>+</sup> cytotoxic T cell. CD4 glycoprotein is expressed on the surface of helper T cells and binds to MHC class II, whereas CD8 glycoprotein is expressed on the surface of cytotoxic T cells and binds to MHC class I (Figure 1.2). Thus, CD4 and CD8 contribute to T cell recognition by helping to focus the cell on MHC presented molecules. For instance, CD4<sup>+</sup> helper T cells recognize antigens displayed on dendritic cells, macrophages, and B cells whereas cytotoxic (CD8<sup>+</sup>) cells recognize any nucleated host cell displaying a foreign peptide with a MHC class I. T Helper (Th) cells secrete effector molecules,

including cytokines that signal to other cells to initiate a specific functional response, including B cells, to secrete antibody to assist macrophages in killing pathogens or pathogen infected cells (Parker, 1993). CD8<sup>+</sup> cytotoxic T cells can undergo clonal expansion to grow an army of effector cells in proportion to the amount of the pathogen recognized and directly kill the infected cell (Huang et al., 2019, Martin and Badovinac, 2018). The cytotoxic activity is activated when a foreign antigen is encountered, which can then induce the secretion of perforin and different types of granzyme (Nicolet et al., 2020, Weigelin et al., 2021). Secreted proteins will then puncture the infected cells and cause cell death (Yamashita et al., 1998, Gricks and Gribben, 2003). Natural killer (NK) cells can also kill infected or damaged cells and in this way are similar to CD8<sup>+</sup> cytotoxic T cells, but they have receptors that are much less specific. Natural killer cells serve to contain viral infections while the adaptive immune response is generating antigen-specific cytotoxic T cells that can clear the infection.

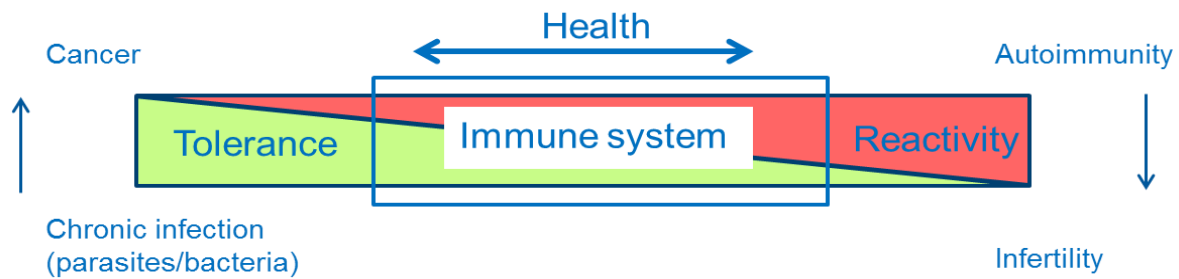


**Figure 1.2: Antigen presentation stimulates CD8<sup>+</sup> cytotoxic cells or CD4<sup>+</sup> helper T cells to mature**

CD4<sup>+</sup> and CD8<sup>+</sup> T cells bind to antigens presented by MHC molecules on the surface of antigen presenting cells. These antigens are typically short chains of amino acids (peptides) derived from proteins and are detected by the T-cell receptor (TCR).

### 1.3 Immune Homeostasis and Tolerance

A healthy immune system requires a fine balance between tolerance and reactivity. Pathology occurs when the balance is disturbed, whereby too much tolerance could potentially lead to cancer or chronic infection, and too much reactivity could lead to autoimmunity (Figure 1.3). Immunological tolerance is the selective ability of the immune system to be unresponsive towards self-antigens, food antigens and harmless microbes in the gut and yet retain the ability to respond to harmful foreign pathogens. Tolerance can be classified into central tolerance, which occurs during the maturation of T cell and B cell populations, and peripheral tolerance which takes place in other tissues and lymph nodes in the periphery. Central tolerance refers to the tolerance established by deleting autoreactive lymphocyte clones before they develop into fully immunocompetent cells. It occurs during lymphocyte development in the thymus and bone marrow for T and B cells, where cells are exposed to a wide array of self-antigens essential for the negative selection of self-reactive cells and the establishment of central tolerance. Those self-reactive cells then undergo apoptosis or become functionally inactive (Yamano et al., 2015). Some self-reactive T cells manage to escape negative selection in the thymus, and therefore peripheral tolerance comes into place to prevent auto-reactivity. Peripheral tolerance is crucial for preventing responses against harmless neo antigens such as food and beneficial microbes in the intestine, which are not present in the thymus during T cell development. A suppressive type of T cell known as a regulatory T cell (Treg) plays a pivotal role in mediating peripheral tolerance and establishes the local tolerogenic environment in the periphery (Sakaguchi et al., 2008). The majority of Tregs are produced in the thymus as a functionally distinct T cell subpopulation, with the T cell receptor (TCR) repertoire considerably skewed toward recognizing self-antigens which play an important role in the periphery to regulate T cell autoreactivity (Sakaguchi, 2003).



**Figure 1.3: Schematic illustrates a fine balance between tolerance and reactivity of healthy immune system**

A healthy immune system requires a balance between tolerance and reactivity by immune cells. When that balance goes wrong, pathology occurs.

## 1.4 CD4<sup>+</sup> T cell Lineage(s) Differentiation

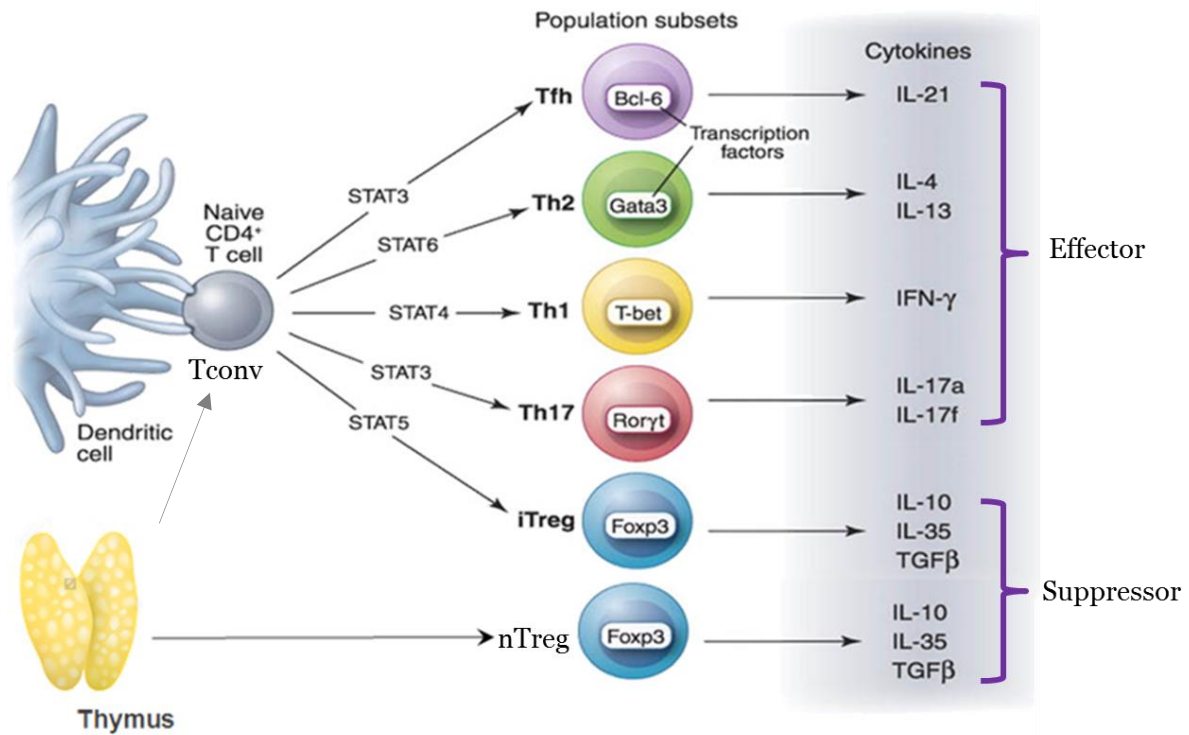
CD4<sup>+</sup> T cells are a heterogeneous population of cells which maintain immune homeostasis in the adaptive immune system. CD4<sup>+</sup> T cells can be separated into two arms based on their functions: the effector arm and the suppressor arm. Those CD4<sup>+</sup> T cells that are involved in the effector arm are responsible for recognising and targeting of foreign pathogens in the body. In order to respond swiftly and specifically to the vast repertoire of immune challenges the human body encounters, specialised sets of T cells can be rapidly differentiated and deployed. These different sets of T cells express specific lineage defining transcription factors (TF) and these TFs orchestrate the expression of effector molecules and cytokines specific for a particular pathogen (Figure 1.4).

T helper type 1 (Th1) cells express T-bet as their master transcription factor and this orchestrates the expression of a suite of effector molecules including IFN $\gamma$ , the main effector cytokine secreted by Th1 cells. This allows these cells to specifically eliminate intracellular (viral, bacterial, and parasitic) infections. The Th1 cells are also crucial for efficacious vaccine function by priming an antibody response. T helper type 2 (Th2) cells express GATA3 as their master transcription factor and this orchestrates the expression of a suite of effector molecules including IL-4, IL-5, IL-13 to expel large, extracellular parasites and as such, protect against ectoparasites and gastrointestinal worms. Th2 cells are also important for the induction and

development of humoral immune responses. IL-4 and IL-13 activate B-cell proliferation, Ig class-switching, and antibody production as well as initiating or promote allergic responses (Mosmann et al., 1986). Likewise, T helper type 17 (Th17) cells that express the transcription factors ROR $\gamma$ t and ROR $\alpha$  are able to produce IL-17 to mediate responses to extracellular bacteria and fungi (Aggarwal et al., 2003). The recently described IL-22-producing Th22 cells are targeted to the skin and may contribute to skin homeostasis and also inflammation (Trifari et al., 2009). Follicular T helper (Tfh) cells that express the transcription factor BCL-6 and produce IL-21, assist B cells in the formation of the germinal centres, which is required for antibody affinity maturation as well as conversion to plasma cells for production of high affinity antibodies (Schaerli et al., 2000).

CD4<sup>+</sup> T helper cell responses need to be curbed to avoid over exuberant and inappropriate responses against exogenous antigens and also when the pathogen is cleared. Several mechanisms have evolved to attenuate these responses to avoid unwanted tissue destruction, immunopathology, and ultimately autoimmunity. CD4<sup>+</sup> CD25<sup>+</sup> regulatory T (Treg) cell which forms the major part of the suppressor arm of the CD4<sup>+</sup> T cell heterogeneous population. They are shaped by the expression of transcription factor FOXP3 and this defines the Treg suppressive function.





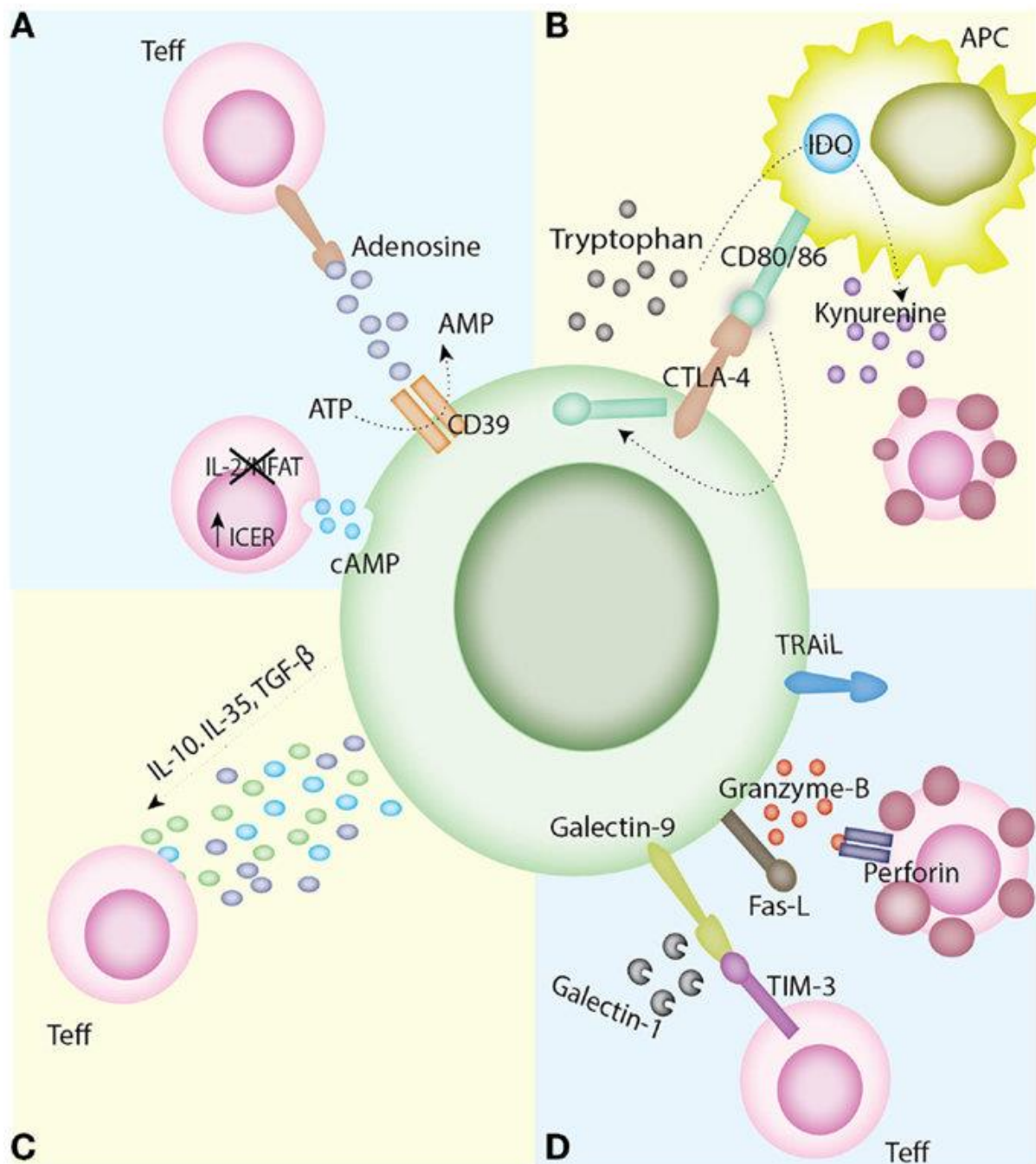
**Figure 1.4: CD4<sup>+</sup> T cell subset differentiation**

CD4<sup>+</sup> T cells show remarkable plasticity and can differentiate into many different subsets based on the cytokines secreted during priming of the subsets by antigen presenting cells (APC). Tfh, Th2, Th1, Th17 are part of the effector arm of the CD4<sup>+</sup> T cell population whereas iTreg and nTreg are part of the suppressor arm of the CD4<sup>+</sup> T cell population. Figure modified from O'Shea and Paul. 2010. Science.

## 1.5 Mechanisms of Treg Action

Tregs are characterized by their ability to inhibit T cell proliferation *in vitro*, as shown in Figure 1.5, suppressing inappropriate immune responses against harmless antigens. Treg carry out their suppressive activities in several ways: Treg disrupt Tconv metabolic pathways by high-affinity CD25 (also known as IL-2 receptor alpha)-dependent cytokine-deprivation-mediated apoptosis and expression of the ectoenzymes CD39 and CD73 (Antonioli et al., 2013, Allard et al., 2017, Feng et al., 2011), which allows adenosine triphosphate (ATP) to be metabolised to adenosine monophosphate (AMP) resulting in the secretion of the immunoregulatory purine, adenosine. Tregs have also been found to express high levels of intracellular cyclic AMP (cAMP). This is transferred to T effector cells through gap junctions, which leads to the upregulation of Inducible cAMP Early Repressor (ICER) and in turn the inhibition of Nuclear Factor of Activated T-cells (NFAT) and IL-2 transcription, resulting in apoptosis by IL-2 deprivation (Deaglio et al., 2007, Fletcher et al., 2009). The second mechanism of suppression by Treg involves the modulation of APC maturation and function, whereby the interaction of cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) on Tregs with its ligand CD80/86 on APCs, delivers a negative signal blocking T cell activation. CTLA-4 acts on APC using a number of different mechanisms, and this includes capturing APC-expressed ligands and subsequent trans-endocytosis and also the upregulation of Indoleamine 2,3-dioxygenase (IDO) and the generation of kynurenines resulting in the suppression of T cells (Qureshi et al., 2011, Read et al., 2000, Takahashi et al., 2000). The third mechanism of suppression occurs if Treg secrete anti-inflammatory cytokines including IL-10, IL-35, and transforming growth factor beta (TGF $\beta$ ), leading to the suppression of T cell activation *in vivo* (Collison et al., 2007, Fahlén et al., 2005, Hara et al., 2001, Powrie et al., 1996). The fourth major mechanism by which Treg suppress an immune response is to directly induce apoptosis via granzyme A/B and perforin, TRAIL, the Fas/Fas-ligand pathway, the galectin-9/TIM-3 pathway, or the production of

galectin-1 (Gondek et al., 2005, Grossman et al., 2004, Allard et al., 2017, Volpe et al., 2016, Strauss et al., 2009, Wang et al., 2009).



**Figure 1.5: Mechanisms of Treg suppression**

A) Disruption of metabolic pathways. B) Modulation of APC maturation and function. C) Anti-inflammatory cytokine production. D) Induction of apoptosis. Image adapted from Safinia et al. (2015).

## 1.6 Thymic Treg and Peripheral Treg

Tregs are now characterized by their site of differentiation, namely thymus derived natural Tregs (tTregs) and peripherally induced Tregs (pTregs). The majority of Treg cells originate from the thymus. These cells are generated when CD4<sup>+</sup> thymocytes encounter a strong interaction with self-antigen. tTregs have high expression of the IL-2 receptor alpha chain (CD25), and stable expression of transcription factor FOXP3. These cells are clearly crucial, since mutations or deletions in FOXP3 (described in detail in 1.8 below) and CD25 genes cause fatal autoimmune diseases in both mice and human (Fuchizawa et al., 2007, Clark et al., 1999, Dominguez-Villar and Hafler, 2018, Eggenhuizen et al., 2020). Treg cells may also differentiate from naïve FOXP3<sup>-ve</sup> CD4<sup>+</sup> T cells in the periphery [called peripherally-derived Treg (pTreg) cells] or *in vitro* after stimulation in the presence of TGF- $\beta$  and IL-2 (termed iTreg cells) (Zheng et al., 2007). In mouse, the pTreg can be identified by co expression of FOXP3 and Neuropilin-1 (Nrp1) (Singh et al., 2015), but to date there is no equivalent biomarker for human pTreg. Studies from mouse models have shown that tTreg cells exhibit strong lineage fidelity, whereas pTreg cells can revert into conventional/effector CD4<sup>+</sup> T cells. The tTreg stronger lineage commitment also makes them the safest cells to use in adoptive cell therapy to treat autoimmune and inflammatory disorders. An absence or impairment of the tTreg population can result in chronic inflammation and autoimmunity. This has been shown in autoimmune diseases such as inflammatory bowel disease, type 1 diabetes, multiple sclerosis and rheumatoid arthritis (Belniak et al., 2007).

The gastrointestinal tract is constantly exposed to food proteins, commensal bacteria, and occasionally pathogenic microorganisms. Despite enormous bacterial challenge, the host intestine establishes a mutualistic relationship with the microbiota, with which it lives in harmony. Among the multiple mechanisms which have evolved to regulate this relationship, Treg cells have been implicated as a dominant element. Experimentally, mice with deficient

Treg cell numbers or Treg that are deficient in the cytokines that would normally mediate Treg cell function, failed to provide protection against chronic intestinal inflammation. However, by adoptively transferring Treg cells back into the mouse intestine, the severity of autoimmune disease can be diminished (Barnes and Powrie, 2009). Activated tTregs can mediate infectious tolerance by delivering TGF $\beta$  to naive Tconv or effector T cells nearby, to generate pTregs in the periphery (Shevach and Thornton, 2014). Peripheral Tregs also participate in the control of immunity at sites of inflammation, especially at mucosal surfaces. FOXP3 expressing human pTreg like cells can also be generated from naïve FOXP3-ve CD4 Tconv cells by stimulating them *in vitro* in the presence of TGF $\beta$ , all trans retinoic acid (ATRA) and rapamycin, and these are known as *in vitro*-induced Tregs (iTregs) (Kim et al., 2020). tTreg and pTreg only account for ~10% of peripheral human CD4+ T cells, but this is sufficient to maintain lifelong immunological tolerance in man. While these cells are an attractive option for immunotherapy to treat human autoimmune diseases, the relative abundance in peripheral blood is low and therefore to achieve the numbers of cells required, these Tregs would need to undergo several rounds of ex vivo stimulation and expansion which can be challenging to achieve. In contrast, iTreg cells, which can easily be generated from large numbers of naïve CD4+ T cell (~50% of peripheral human CD4+ T cells) ex vivo, may offer an alternative to nTreg and pTreg cells. This has enormous implications for immunotherapy, but there are problems associated with the use of iTreg as a therapeutic: iTreg can be unstable both long term and in pro-inflammatory conditions. iTregs may not fully recapitulate the functional or phenotypic characteristics of *in vivo* generated pTregs (Lu et al., 2011). iTreg instability and high plasticity have been demonstrated, showing an easy conversion of the cells to a pro-inflammatory or pathogenic Th17 phenotype in the presence of IL-6, IL-1, and TGF $\beta$  *in vitro* (Yang et al., 2008), resulting in exacerbated and not reduced inflammatory immune responses.

As well as sharing similar suppressive functions with tTreg, both pTreg and iTreg cells have separate non-overlapping phenotypes, including different TCR diversity for detecting, and eliciting responses to, non-self-antigens, particularly those present at the mucosal interface (Haribhai et al., 2011). Helios has been, somewhat controversially, identified as a marker to identify tTreg (Thornton et al., 2010), but has limited usefulness for isolation, since it is a transcription factor and therefore not expressed on the cell surface. Recent studies from Opstelten et al. (2020) have described GPA33 as a useful surface marker for identifying and isolating pure populations of human tTreg that express stable levels of Helios and FOXP3 even after in-vitro culture, but as yet there are no validated biomarkers for isolation of human pTreg from the total Treg population. This contrasts with the ability in mouse, to easily identify and isolate pTreg based on the expression of surface marker Nrp1, distinguishing them from tTreg (Singh et al., 2015). It is still unclear if human iTregs and pTregs are phenotypically and functionally identical populations that use the same suppressive mechanisms, since pTregs have been shown to be a more stable population *in vivo* compared with iTregs (Mikami et al., 2020, Ohkura et al., 2012). Interestingly, epigenetic studies of both mouse and human Foxp3 locus by Floess et al. (2007) have demonstrated complete demethylation within an evolutionary conserved region upstream of exon 1 of the Foxp3 locus, and this region has been named the Treg-specific demethylation region (TSDR). Demethylation of the TSDR region is a specific marker of tTregs, as Tconv cells and *in vitro*-generated iTregs display an almost complete methylation of the TSDR. The status of TSDR methylation may be, therefore, critical for maintaining stable FOXP3 expression and a fully functional Treg phenotype in iTregs.

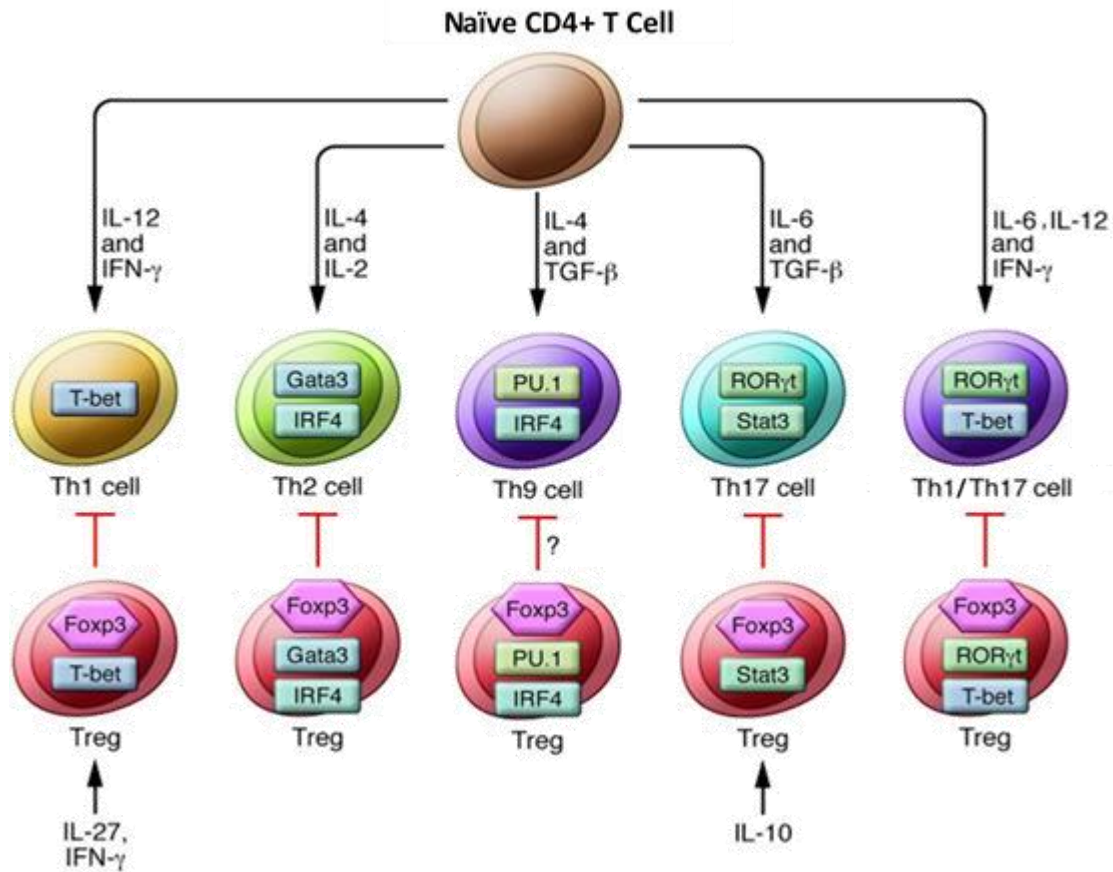
It is now apparent that Treg are far more heterogeneous than previously observed and are adaptable to specific tissue microenvironments (Wing and Sakaguchi, 2012). This phenotypic plasticity allows for swift and specific responses to an effector action, and this is crucial in preventing chronic inflammation and autoimmunity.

## 1.7 Mirrored Suppressive and Effector Responses

The Treg compartment is much more sophisticated than has previously been thought (Duhén et al., 2012). As well as expressing the master transcription factor FOXP3, Treg can express Th-associated transcription factors to orchestrate transcriptional programs to specifically regulate Th1, Th2, Th17, Th22, Th9 and Tfh responses as shown in Figure 1.6 (Höllbacher et al., 2020, Hope et al., 2019, Xie and Dent, 2018). This allows the Treg to specifically follow and counteract an effector immune response. Treg cells also express diverse patterns of chemokine receptors to direct them to sites of inflammation throughout the body in both lymphoid and non-lymphoid tissues (Duhén et al., 2012). Th1-like Treg (Treg1) for example, express FOXP3, the Th1 transcription factor T-bet and the chemokine receptor CXCR3 and these initiate a cascade of gene expression to program the Treg to control type 1 inflammation (Levine et al., 2017, Koch et al., 2009). In experimental mouse models, this has been shown to be crucial for preventing autoimmunity. Treg cells that cannot upregulate T-bet fail to persist and control autoimmunity when transferred into Treg deficient mice (Levine et al., 2017), and additionally, T-bet deficient Treg cannot attenuate IFN $\gamma$ -driven pathology in mice infected with *Toxoplasma gondii* (Hall et al., 2012). Treg subsets expressing CXCR3 and T-bet have also been described in humans (Duhén et al., 2012). Furthermore, studies by Di Giovangiulio et al. (2019) showed that a higher proportion of Tregs from the gut of patients with inflamed Inflammatory Bowel Disease (IBD) co-expressed FOXP3, T-bet and, interestingly, IFN $\gamma$  which was exclusively expressed by T-bet<sup>+</sup> Tregs in mucosal biopsies from patients with Crohn's disease (CD) and ulcerative colitis (UC) compared with healthy controls. These results suggest that Th1-like Tregs (Treg1), characterized by the expression of T-bet, when expressing IFN $\gamma$ , may play a role during inflammatory flares in IBD patients. Emerging evidence from Tan et al. (2016) in a type 1 diabetes mouse model, suggests that T-bet is critical for Treg to control the aggressive infiltration by effector cells in insulinitis. Thus, when T-bet is ablated, Treg cells are unable to

restrain an autoimmune attack on the pancreas. This highlights the importance of the T-bet<sup>+</sup>Treg cells, suggesting they have a non-redundant role that cannot be replaced by other Treg subsets. More recently, T follicular regulatory cells (Tfr) provide a clear example of Tregs acting as “helper” cells for the immune response by producing IL-10 that promotes germinal centre B cell growth and the GC-dependent high-affinity Ab response. Therefore, in the context of the germinal centre response, Tfr cells appear to maintain a key balance between help and suppression of Tfh cell, germinal centre B cell, Tfh cell cytokines, and auto-antibodies (Sage and Sharpe, 2015, Xie and Dent, 2018). It is now apparent that these paired T helper-Treg subsets have the ability to maintain immune homeostasis and prevent autoimmunity, dependent on their appropriate co-localization with their paired effector T cells. However, the degree of phenotypic and functional concordance between different T helper and Treg cell subsets has not been examined carefully and systematically in humans.





**Figure 1.6: Phenotypically mirrored Treg and Tconv helper lineages**

Peripheral naïve CD4<sup>+</sup> T cell differentiation into Th1, Th2, Th17, Tfh and Th9 cell which forms the effector arm of the CD4 T cells population. Each of these T helper lineages have a partnering Treg that express T helper lineage defining transcription factor which forms the suppressor arm of the CD4 T cell population (Chaudhry and Rudensky, 2013)

## 1.8 Molecular Mechanisms of FOXP3 Function

The forkhead (FKH) transcription factor FOXP3 is well documented as the lineage defining transcription factor in Treg (Rudensky, 2011, Sakaguchi, 2000). It has a crucial role in Treg development and the suppressive function of Treg. The importance of FOXP3 was first reported in mice with mutated FOXP3 (Brunkow et al., 2001). These mice have reduced Tregs and are born with a severe autoimmune disorder that leads to destruction of multiple organs and early lethality, known as scurfy. In humans, loss of function of FOXP3 has been associated with a similar disease: Immunodysregulation, Polyendocrinopathy, Enteropathy, X-linked (IPEX) syndrome; a severe autoimmune and inflammatory disorder which also affects multiple organs (Bennett et al., 2001). Such findings highlighted the importance of FOXP3 in the development of a functional Treg cell. However, it has been observed in autoimmune disease cases, that rarely is there loss of function or mutated FOXP3, but more commonly there is disruption to the FOXP3 controlled regulatory network which shapes the Treg suppressive function. FOXP3 is known to regulate a network of genes including transcription factors (TF) and microRNAs (miRs) to enforce the Treg phenotype (Ha, 2011, Rudra et al., 2012, Sadlon et al., 2010).

Genome-wide chromatin immunoprecipitation experiments coupled with mRNA and microRNA expression analysis, has demonstrated that FOXP3 mainly functions as a sequence specific transcriptional repressor (Sadlon et al., 2010). The Barry lab has been interested in identifying the mechanisms of action of FOXP3, as well as the key genes controlled by FOXP3 in order to determine how FOXP3 shapes Treg function and how perturbations to this may be the cause of autoimmune disease. We have focussed on genes that are repressed by FOXP3 as it is clear that this is a means by which effector function in Treg is restrained (Beyer et al., 2011). Additionally, we identified all regions of the human genome where FOXP3 is bound using, chromatin immunoprecipitation, as a means of identifying key genes regulated by FOXP3 (Sadlon et al., 2010). FOXP3 targets included genes expressing cell surface molecules,

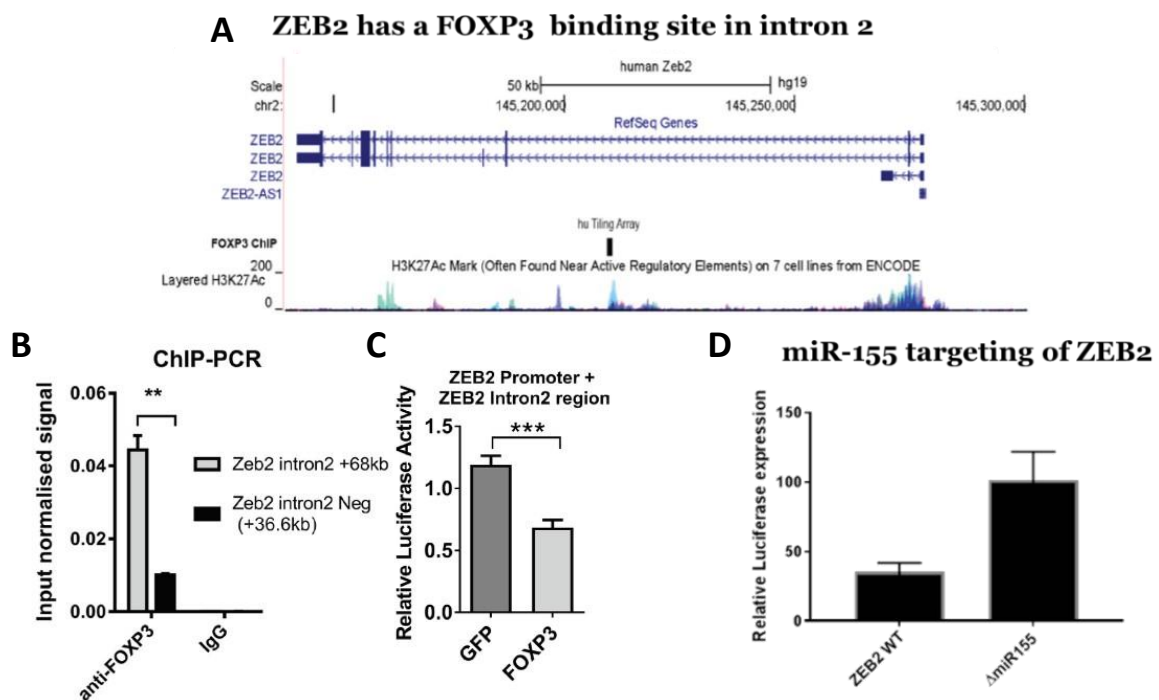
cytokines, and microRNAs (miRs). When we overlapped the FOXP3 target datasets with genes that are differentially regulated in Treg compared with Tconv, the result showed 739 genes, some of which are microRNAs.

Using a prioritization strategy, we selected genes that are both a target of a FOXP3-induced miR and a FOXP3 target, and ranked on functional criteria such as the target being a transcription factor. This approach has allowed us to identify a number of genes of interest, including SATB1 and also ZEB2 (Beyer et al., 2011, Brown et al., 2018). The Barry lab has shown that repression of SATB1 by FOXP3 in Treg is required for Treg suppressive function. We have shown that FOXP3 and FOXP3 induced miR-155 repress SATB1 and ZEB2 in breast cancer cells, demonstrating the regulatory loops for tight repression of SATB1 and ZEB2 in the presence of FOXP3 and FOXP3-induced miR-155 (McInnes et al., 2012, Brown et al., 2018). Since ZEB2 is regulated by FOXP3 and FOXP3-induced miR-155 in breast cancer and is tightly repressed in human Treg, where FOXP3 and FOXP3-induced miR-155 are highly expressed, this suggests that ZEB2 is specifically repressed in a human Treg, most likely by FOXP3 and miR-155, to potentially prevent an unfavourable phenotype.

## **1.9 ZEB2 – A Key Gene Repressed by FOXP3 in Treg cells**

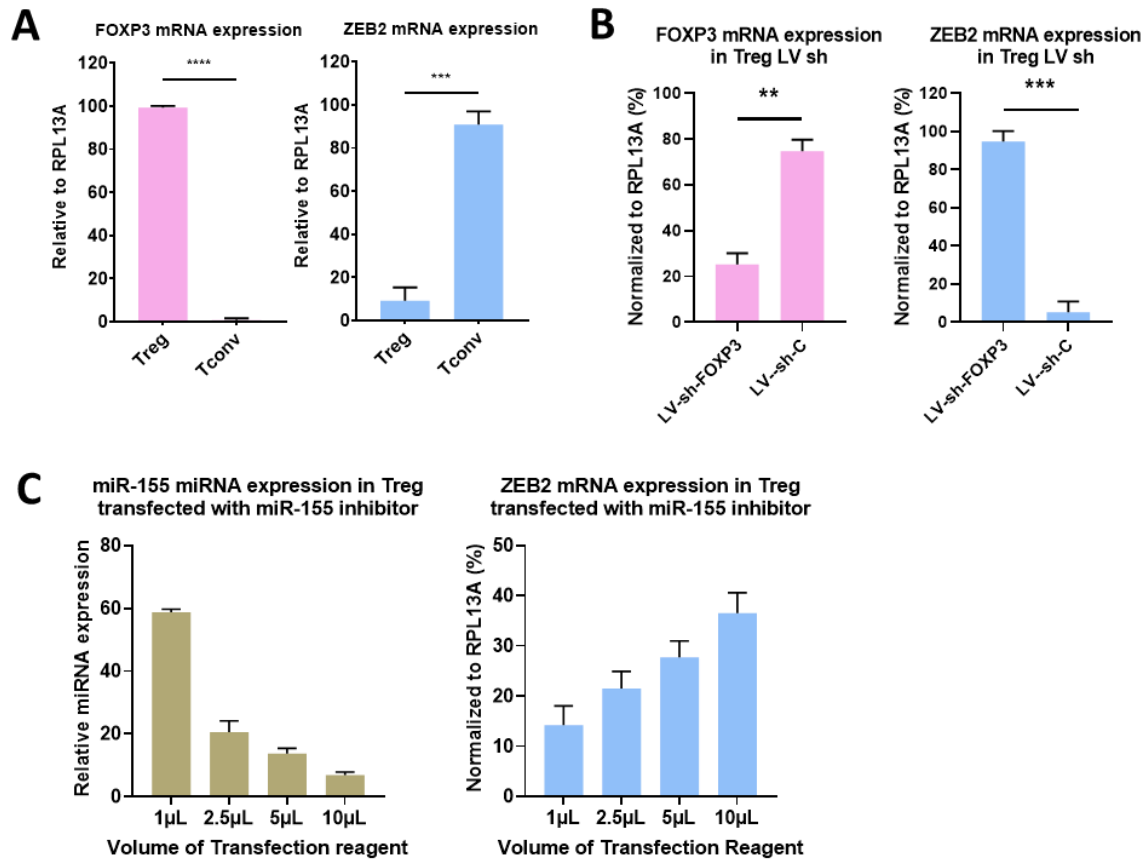
ZEB2 has been identified as one of the genes that is differentially regulated by FOXP3 in Treg compared with Th cells. In this study, a single FOXP3 binding region was detected by ChIP-ChIP downstream of the transcription start site in intron 2 (Sadlon et al., 2010). This was confirmed with ChIP PCR and then functionally validated using luciferase-reporter assays, in which FOXP3 inhibited luciferase activity in reporter vectors containing the promoter and the identified FOXP3 binding region in intron 2 of the ZEB2 gene (Brown et al., 2018). Moreover, using ZEB2 3'UTR truncations to selectively remove each of the 4 miR-155 target sequences followed by miR-155 target sequence mutations, we identified, using luciferase assays, the specific targeting of ZEB2 by miR-155. In this paper we characterised the repression of ZEB2

by FOXP3 and miR-155 and their role in preventing breast cancer progression by tightly suppressing ZEB2 expression (Figure 1.7). Since these mechanisms also promote tight repression of ZEB2 in Treg, as shown in Figure 1.8, we can speculate that ZEB2 inhibition is required in Treg potentially to enforce a functional suppressor Treg phenotype (Wong & Brown et al., manuscript in preparation). In contrast, in the absence of FOXP3-mediated repression in Th cells, ZEB2 is highly expressed, although the role of ZEB2 in these cells is still uncharacterised. One important question is; in which compartment of T helper lineage is ZEB2 expressed and why?



**Figure 1.7: ZEB2 is a bona fide target of FOXP3 and miR-155**

A) ZEB2 has a FOXP3 binding site in intron 2 with ChIP-chip data. B) and C) Confirms FOXP3 binds to the region at intron 2 with ChIP-PCR and show repressive activity by FOXP3 in a Luciferase-reporter assay. D) ZEB2 is a target of miR-155 at 3'UTR. Figure adapted from (Brown et al., 2018).

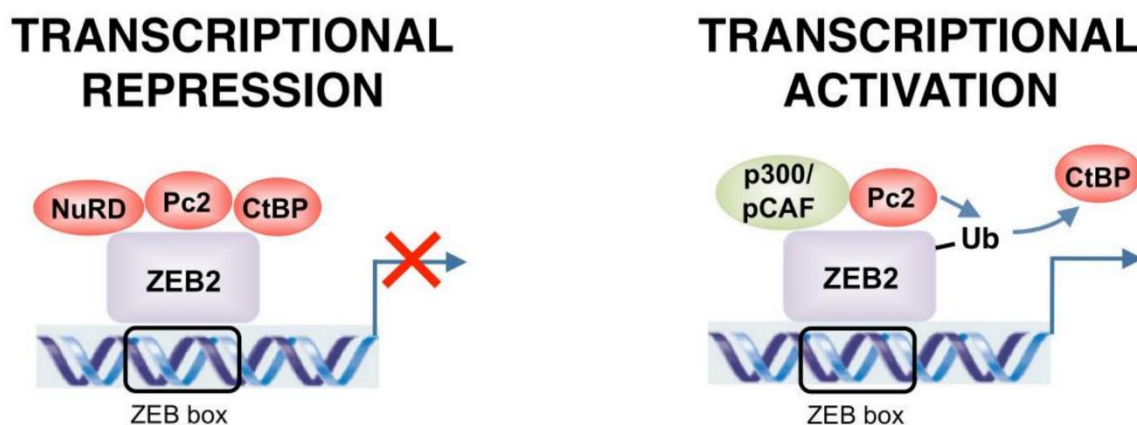


**Figure 1.8: Expression and regulation of ZEB2 in Treg and Tconv by FOXP3 and miR-155**

A) ZEB2 (left) and FOXP3 (right) expression in Treg and Tconv. Relative abundance of ZEB2 and FOXP3 mRNA was quantitated by qRT-PCR using the standard curve method for relative quantitation and expressed relative to reference gene RPL13A mean + SEM. ZEB2 expression,  $n = 4$  experiments,  $***p < 0.001$ . FOXP3 expression,  $n = 8$  experiments,  $****p < 0.0001$ . B) ZEB2 (left) and FOXP3 (right) expression in Treg where FOXP3 knocked down with LV-sh-FOXP3 or transduced with control LV-shC. Relative abundance of ZEB2 and FOXP3 mRNA was quantitated the same as (A) with mean + SEM. ZEB2 and FOXP3 expression,  $n = 3$  experiments,  $**p < 0.01$ ,  $***p < 0.001$ . (C) Treg transfected with 1μL, 2.5μL, 5μL or 10μL of miR155 inhibitor, left showing mir155 expression and right ZEB2 expression. Relative abundance of miR-155 miRNA was normalised to reference RNU-24.

## 1.10 Molecular Structure and Transcriptional Activities of ZEB2

ZEB2 is also known as SIP1, KIA0569, SMADIP-1 and Zfhx1b (Amiel et al., 2001, Postigo and Dean, 2000). It is a two-handed zinc-finger transcription factor that directly binds the DNA sequence 5'-CACCT-3' (Verschueren et al., 1999) in several promoters, and it can also interact with other transcription factors and cofactors. ZEB2 may also be a direct competitor of other enhancer box-protein (E-protein) including its close relative, ZEB1 (structurally similar to ZEB2) and ID2, for E-box binding sites, owing to similarity in consensus sequences (Sekido et al., 1994). ZEB2 can also mediate transcriptional repression via cooperation with activated Smads or through recruitment of the corepressor C-terminal binding protein (CtBP), Pc2 as well as histone deacetylase complexes, particularly NuRD (van Grunsven et al., 2007, Verschueren et al., 1999, Verstappen et al., 2008, Postigo and Dean, 2000). Activation of transcription by ZEB2 also involves recruitment of coactivators such as p300 and pCAF as well as sumoylation of ZEB2 by polycomb Pc2 which disrupt interaction of CtBP with ZEB2 (Figure 1.9).



**Figure 1.9 ZEB2 could act as either transcriptional repressors or activators depending on the target gene and tissue**

Transcriptional repression and activation is achieved through differential recruitment of cofactors. Post-transcriptional modifications of ZEB2 alter the set of coactivators and corepressors bound and switch both proteins from transcriptional repressors to activators. Figure adapted from Sánchez-Tilló et al. (2011).

## 1.11 Clinical Significance of ZEB2

ZEB2 plays a crucial role in the early development of neural crest cells during foetal development, which involves many tissues as well as specific organs in the body. Mutations in the ZEB2 gene are associated with the Mowat–Wilson syndrome (Mowat et al., 2003). This disease exhibits mutations and even complete deletions of the ZEB2 gene, leading to altered or non-functional ZEB2 which may culminate in abnormal development of multiple organs. Many of the symptoms can be explained by the irregular development of the structures from the neural crest (Dastot-Le Moal et al., 2007). In Hirschsprung's disease, loss of ZEB2 results in abnormal nerve development in the digestive tract, resulting in severe constipation and enlargement of the colon (Saunders et al., 2009). The mechanism of action of ZEB2 in the enteric nervous system has yet to be determined, but since it is a transcription factor it is highly likely to coordinate a program of development by directing the transcriptional regulation of enteric-associated genes.

ZEB2 and its highly conserved homologue, ZEB1, also play a role in promoting cancer metastasis of breast, ovarian, colorectal and prostate cancer (Fardi et al., 2019). These ZEB proteins are well-known to induce epithelial to mesenchymal transition (EMT), a process that reorganizes epithelial cells to become migratory mesenchymal cells. Epithelial gene E-cadherin (CDH1) and mesenchymal gene Vimentin are known target genes of ZEB1 and ZEB2, and downregulation and upregulation of these gene respectively is considered a hallmark of EMT (Wang and Zhou, 2013). Vimentin and E-cadherin are commonly known to have a normal biological role in early embryonic development (Goossens et al., 2011, Li et al., 2016b), and wound healing (Yin et al., 2013). The miR-200 family have been shown to suppress ZEB1 and ZEB2 and these microRNAs are significantly down-regulated in TGF $\beta$ -induced mesenchymal cells and cancer cells with mesenchymal characteristics (Gregory et al., 2008). Importantly, we have identified an independent regulatory network for controlling ZEB2 alone, separating the

regulation of ZEB1 and ZEB2 and this has allowed us to identify a specific role for ZEB2 independently of ZEB1. We have determined that ZEB2, but not ZEB1, is regulated by FOXP3 and FOXP3-induced miR-155, and this has revealed a non-redundant role for ZEB2, separate from that of ZEB1, in controlling Vimentin expression and, interestingly, not E-cadherin expression in human breast cancer cells (Brown et al., 2018).

## 1.12 ZEB2 Regulates Functional Maturation in Immune Cells

The role of ZEB2 is well established in multiple cancers, where it is known to promote cancer progression and EMT. However, when the Barry lab discovered that it is regulated by FOXP3 and FOXP3 induced microRNA-155 (miR-155) in breast cancer, since FOXP3 and miR-155 are crucial for Treg formation and function, this suggested a potential role in immune cells as well. Our ChIP-ChIP and exon array data indicated that ZEB2 is a target of FOXP3 and is differentially expressed in Treg and Tconv. We then showed that it is kept tightly repressed by FOXP3 and miR-155 in Treg but its expression is high in Tconv where FOXP3 and miR-155 are low. This data suggested that ZEB2 has a role in Tconv, and although we had no information at the time as to what its role might be, several recent lines of evidence suggest that ZEB2 might promote effector actions, since it was emerging in the literature that ZEB2 has a role in several key cell types of the immune system.

Previous studies have shown that ZEB2 is crucial for mammalian embryonic development (Goossens et al., 2011). Moreover, loss of ZEB2 specifically in the murine adult hematopoietic system, resulted (after 10–12 months) in splenomegaly with enlarged spleens containing a significant infiltration of hematopoietic stem cells and megakaryocyte/erythroid progenitor cells, indicative of extra medullary haematopoiesis (Li et al., 2016b, Wu et al., 2016). Initially, ZEB2 was thought to be expressed in B cells but not in T cells (Postigo and Dean, 2000), but more recent findings indicate a role for ZEB2 in several different cells of the immune system. ZEB2 plays a role in the maturation of immune cells including CD8<sup>+</sup> T cells, natural killer



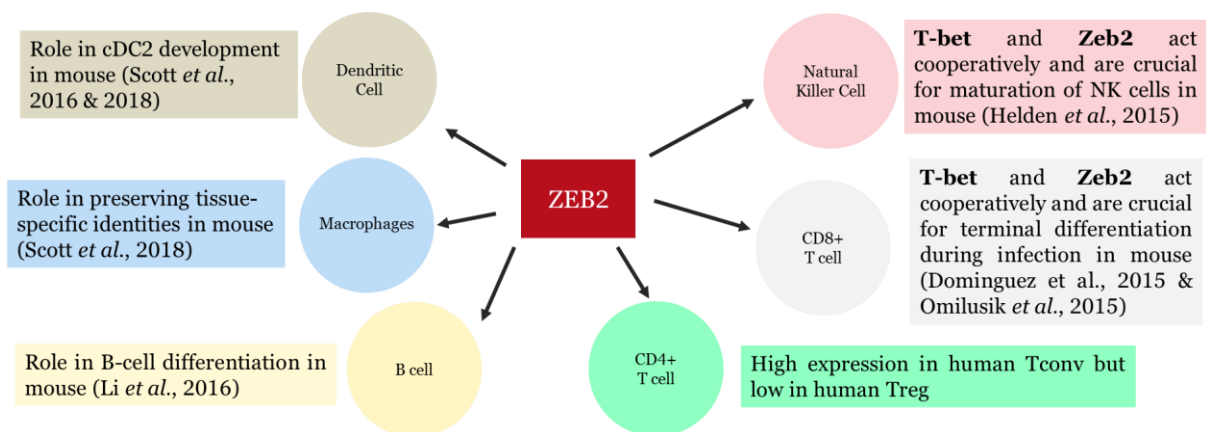
cells, B cells, macrophages and dendritic cells (Li et al., 2016a, Scott et al., 2016, Scott et al., 2018, Wu et al., 2016, van Helden et al., 2015, Omilusik et al., 2015, Dominguez et al., 2015).

Deletion of ZEB2 in mouse B cells, results in the inability of B cells to mature into antibody producing plasma cells (Li et al., 2016a). Deletion of ZEB2 from macrophages results in loss of macrophage tissue residency likely through necroptosis (Scott et al., 2018). ZEB2 has also been shown to be critical for the functional maturation of dendritic cells (Scott et al., 2016, Wu et al., 2016). The expression of ZEB2 in mouse dendritic cells favours the differentiation program into conventional dendritic cell 2 (cDC2) over conventional dendritic cell 1 (cDC1). cDC2 presents antigen only to CD4<sup>+</sup> T cells whereas cDC1 presents antigen only to CD8<sup>+</sup> T cells (Schlitzer et al., 2015). In mouse CD8<sup>+</sup> T and Natural Killer (NK) cells, T-bet, a lineage defining transcription factor for the Th1 T helper subset, induces ZEB2, resulting in the direct and cooperative regulation of functional maturation of these cells (van Helden et al., 2015, Dominguez et al., 2015, Omilusik et al., 2015).

ZEB2 is important in mouse CD8<sup>+</sup> T cell clonal expansion where it promotes the expression of effector genes. When ZEB2 is deleted, CD8<sup>+</sup> T cells are unable to terminally differentiate and carry out cytolytic activity in response to infection (Omilusik et al., 2015, Dominguez et al., 2015). Interestingly, although the miR-200 family of miRNAs have been shown to negatively regulate both ZEB1 and ZEB2 in the context of epithelial differentiation, only ZEB2 mRNA appears to interact with, and be efficiently targeted by, the miR-200 family miRNAs in CD8<sup>+</sup> T cells (Gagnon and Ansel, 2019). ZEB2 expression is also required for functional maturation of NK cells. The loss of ZEB2 in mouse NK cells affects their maturation, survival and function, since they are no longer capable of exiting the bone marrow to elicit their function (van Helden et al., 2015). Furthermore, Goossens et al. (2015) and (Li et al., 2016a) showed that ZEB2 drives immature T cell lymphoblastic leukaemia (TLL) and acute myeloid leukaemia (AML)

respectively in mice. However, there is still little known about the expression and role of ZEB2 in human CD4+ T cells (Figure 1.10).

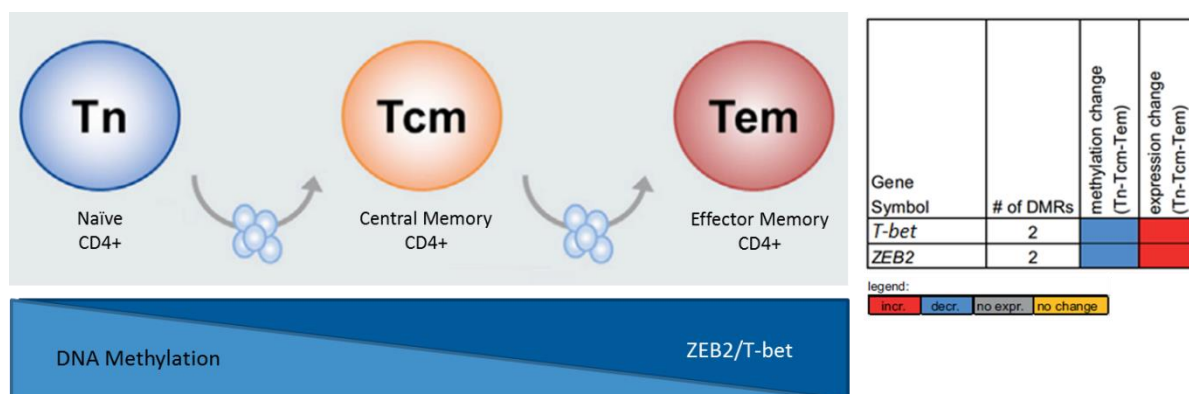
We have shown that ZEB2 is expressed more highly in human T helper cells compared with human Treg (Sadlon et al., 2010), however, whether ZEB2 has a similar role in promoting functional maturation from naïve Tconv to a specific lineage of T helper among the CD4 T cell pool is yet to be elucidated.



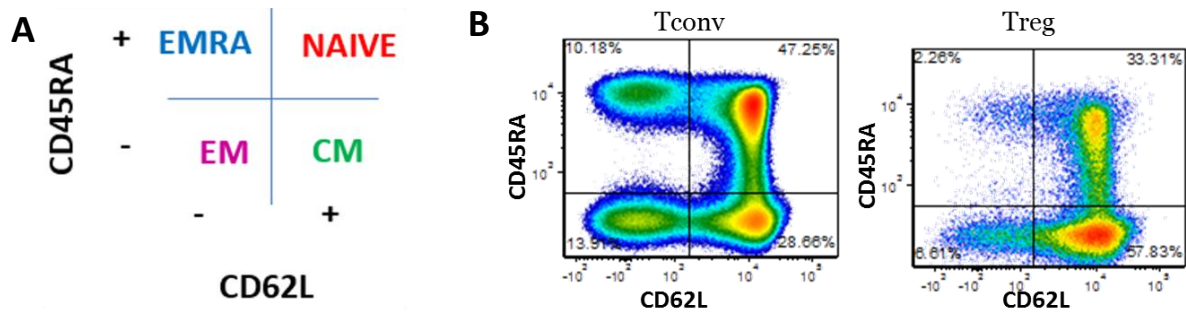
**Figure 1.10: Summary of ZEB2 roles in DC, macrophages, B-cell, CD8+ T cell, NK cell and CD4+ T cell**

### 1.13 ZEB2 is expressed in Th1 Cells

A recent study from Durek et al. (2016) has shown that as human CD4<sup>+</sup> cells progress from naïve, central memory to effector memory, DNA methylation decreases (Figure 1.11) and ZEB2 and T-bet expression increases accordingly. Using two surface markers (CD45RA and CD62L), CD4<sup>+</sup> T cells can be separated into several memory subtypes: naïve, central memory, effector memory and terminal effector memory (effector memory RA<sup>+</sup>) (Sallusto et al., 2004) (Figure 1.12). Naïve cells are quiescent and have not been exposed to antigen within the periphery. Central memory cells are known to circulate but can migrate towards secondary lymphoid organs. Effector memory cells are present in the peripheral tissue ready to activate into effector cells to clear pathogens. Durek et al. (2016) showed that in the effector memory of CD4<sup>+</sup> T cells, where the chromatin state of the effector phenotype cells was more accessible to transcription factor regulation, there was high expression of ZEB2 and T-bet. This study only investigated ZEB2 and T-bet in the effector memory pool of the CD4<sup>+</sup> T cell population, hence further detailed characterization was required to identify the Treg and T helper subsets within the pool population that express ZEB2 and T-bet.



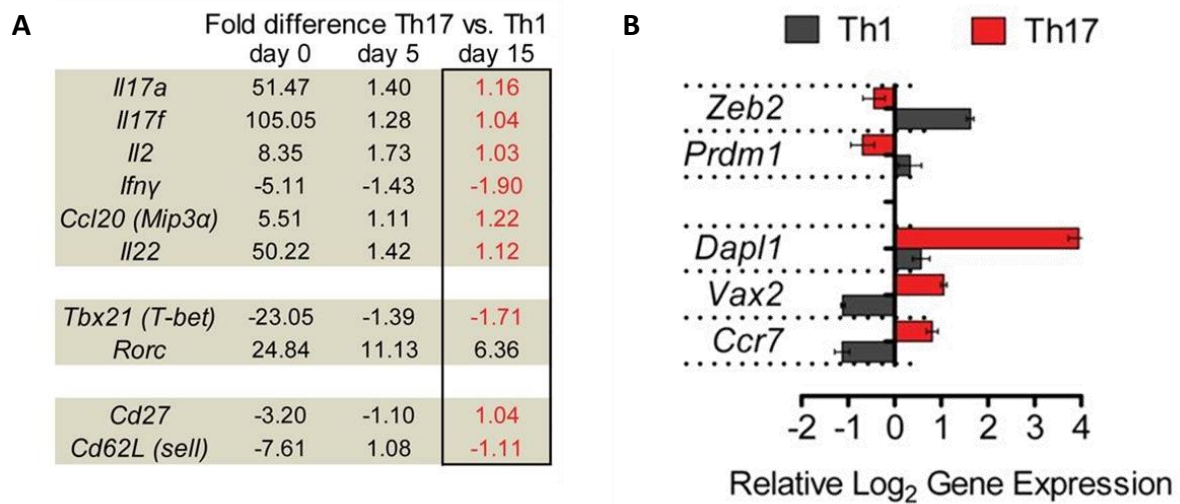
**Figure 1.11: ZEB2 and T-bet expression is high in the effector memory of CD4<sup>+</sup> T cell**



**Figure 1.12: CD4<sup>+</sup> T cell memory phenotype panel**

A) Representative memory phenotype panel where the T-cell population can be separated into 4 memory populations: Naïve, Central Memory (CM), Effector Memory (EM) and Effector Memory RA+ (EMRA). B) Treg and Tconv population segregated into different memory compartments based on CD45RA and CD62L surface markers. Figure adapted from Hope et al. (2019).

T-bet and ZEB2 are more highly expressed in mouse Th1 compared with Th17, providing some insight into ZEB2 expression in 2 subsets of the effector arm of the murine CD4<sup>+</sup> T cell (Muranski et al., 2011) (Figure 1.13), but again Treg were not investigated. Thus, a detailed investigation of ZEB2 expression has yet to be carried out in human Treg and in all of the T helper subsets. Given the published data that T-bet coupled with ZEB2 are required for effector Th1 differentiation, one unanswered question is whether human ZEB2 expression is restricted to the Th1 lineage. The role of Th1 in inflammation and immuno-regulation is to release IFN $\gamma$  (Mosmann et al., 1986). IFN $\gamma$  induces the presentation of antigen by antigen presenting cells, which stimulate IgG2a production by B cells (Peng et al., 2002), and induces the expression of other cytokines and chemokines required for the recruitment of myeloid cells such as macrophages to the site of inflammation (Iriguchi et al., 2015). Th1 cell migration is likewise orchestrated by T-bet, which directly controls the expression of the chemokine receptor CXCR3 and the chemokines CCL4 and CCL3 (Fu et al., 2015). Since ZEB2 has a well-established role in directing tumour metastasis, via promoting migration and invasion of cancer cells, we could hypothesise from this that ZEB2 may be involved in the migration of Th1 cells.



**Figure 1.13: ZEB2 and T-bet expression is higher in Th1 compared with Th17 in mouse**  
 A) Fold differences in T-bet expression comparing cells Th17 culture conditions vs Th1 culture conditions from day 0 to day 15. B) ZEB2 expression is lower in Th17 compared with Th1. Figure adapted from Muranski et al. (2011).

## 1.14 The Unanswered question: Why is ZEB2 in Expressed in human CD4+ T Effector Cells?

ZEB2 has a major role in promoting cancer metastasis in a number of cancers including breast cancer. The Barry lab has shown that in breast cancer its expression can be controlled both transcriptionally by FOXP3 and post-transcriptionally by miR-155, but FOXP3 has a much more established/accepted role in the immune system. FOXP3 is essential for correct Treg formation and function and FOXP3-induced microRNA, miR-155 is likewise essential for correct Treg function. Our finding, therefore, that ZEB2 is part of the FOXP3/miR regulatory network in Treg, as well as in breast cancer, is very interesting. It suggests that ZEB2 is repressed for a reason. When we discovered that ZEB2 is highly expressed in T effector cells (Tconv) this suggested that the role of ZEB2 might be to initiate or promote a T effector function which must be repressed in a Treg in order to maintain immune homeostasis. This has enormous implications for immune function and potential for therapeutic intervention.

Based on the roles of ZEB2 in other immune compartments, it is plausible that ZEB2 may promote effector function in CD4+ T cells. The rationale for this that in natural killer cells (NK) and CD8+ T cell, ZEB2 is shown to be directly induced by T-bet which is also a lineage defining transcription factor for the effector function and differentiation of CD4+ Th1 cell subset. Although it was shown that ZEB2 promotes effector function in other immune cells in the presence of T-bet, it is possible that ZEB2 shares a similar and non-redundant role with T-bet, as shown in Figure 1.10. Dominguez et al. (2015) and Omilusik et al. (2015) showed that ZEB2 represses memory precursor genes and induces effector genes that enhance cytotoxic activity of CD8+ T cell during a recurring infection. This could suggest that T helper cells expressing ZEB2 might have a better recall response when re-encountering the same antigen. We have shown that ZEB2 is expressed at higher levels in human Th as compared with human Treg and is repressed by FOXP3 in Treg. It is likely therefore, that ZEB2 has a similar role in promoting

functional maturation either in a specific lineage of human T helper (presumably Th1 cell), or in all lineages. In a Treg, FOXP3 suppresses the effector genes that are normally regulated by the T helper lineage defining transcription factor, to ensure that the Treg retains its suppressive phenotype. If ZEB2 is an effector gene, as suggested in various immune cell compartments, it might still be repressed by FOXP3 even in a T-bet<sup>+</sup> Treg. These pieces of data all point to the possibility that ZEB2 has a critical role in Th1 effector responses and is repressed in Treg.

A healthy immune system requires specific responses from specific cells and if the effector molecules are not produced correctly or if the lineages are not formed, then this will result in an unbalanced or inappropriate immune response leading to inefficient pathogen removal, or autoimmunity.

My PhD project is to examine how ZEB2 is involved in the effector response of a CD4<sup>+</sup> T cell. Will the dysregulation of ZEB2 in T effector and Treg lead to autoimmune diseases? In order to help determine the role of ZEB2 in T effector cells, first I will characterize its expression pattern across different subsets and populations of CD4<sup>+</sup> T cells. Once I have established where ZEB2 is expressed, this will help me to postulate the type of role ZEB2 may have in CD4 T cells. This could be in the formation of a T cell compartment or in its function. Next, I will determine the network of interactions of ZEB2 manipulating its expression and computational analysis to predict its function and role.

## 1.15 Hypothesis and Aims

Hypothesis: ZEB2 is critical for the function and differentiation of human Th1 cells.

This project aims to complete a comprehensive characterisation of ZEB2 expression in human T helper subsets and Treg subsets and perform functional analysis on these populations to determine the normal biological role of ZEB2 in the immune system and to investigate this in autoimmunity.

Aim 1: Determine the expression of ZEB2 in human in CD4+ T cell subsets.

Aim 2: Manipulation of ZEB2 gene expression to determine its role in CD4+ T cells.

Aim 3: To confirm a link with ZEB2 and autoimmune diseases.



## **CHAPTER 2: MATERIALS & METHODS**

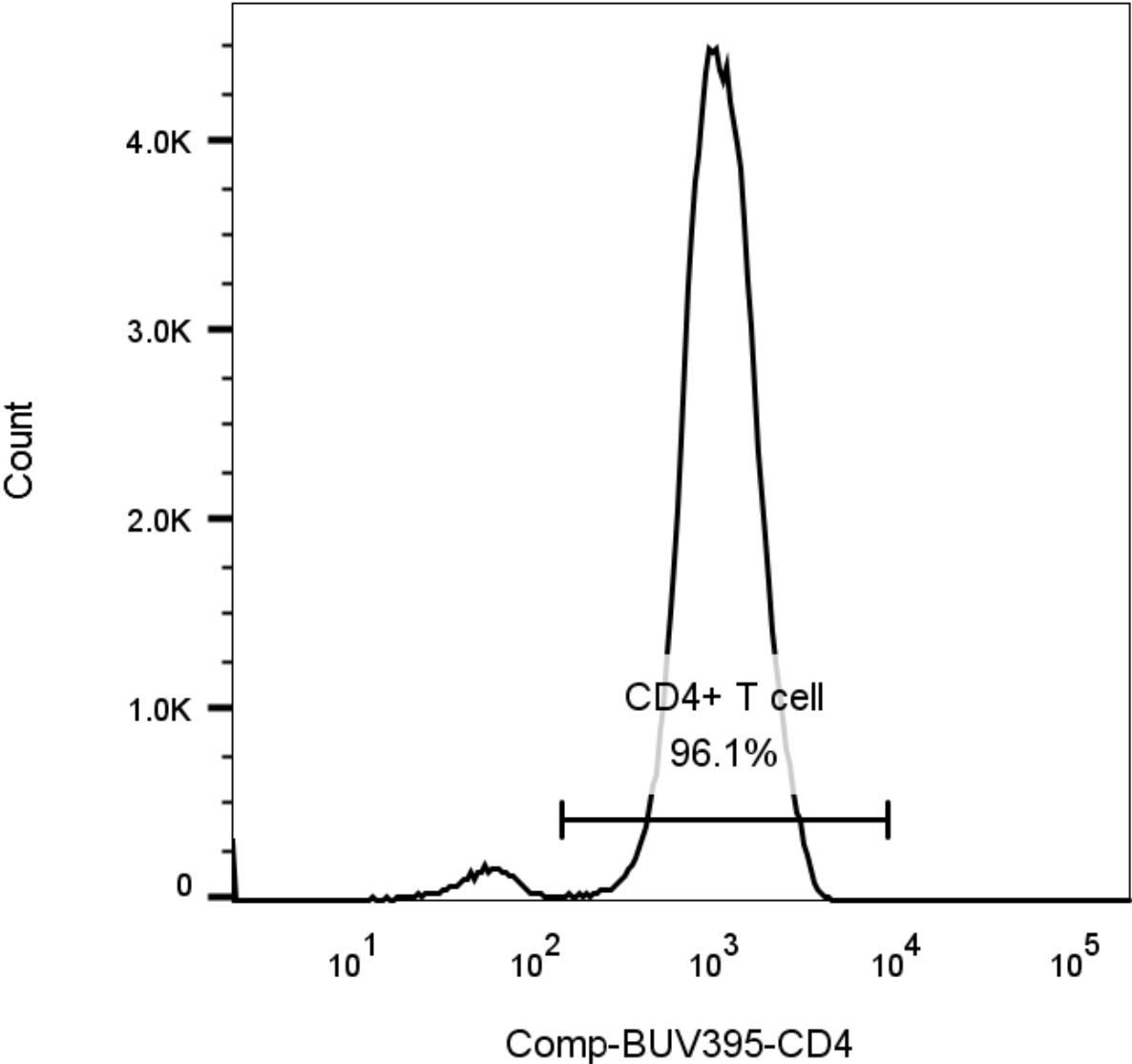
## 2.1 T Cell Isolation, Culture and Antibody Labelling of Proteins

### 2.1.1 Isolation of peripheral blood mononuclear cells

Whole blood (10-20mL) from healthy adult donors was collected in Lithium-Heparin anticoagulant tubes after informed consent (HREC/19/WCHN/65) and buffy coats were obtained from the Australian Red Cross with approval for research use. Buffy coats were used in experiments where larger numbers of cells were required and cell sorting was used to isolate cells of interest. In all other experiments fresh whole blood was used. Peripheral blood mononuclear cells (PBMC) were isolated from whole blood or buffy coats using density gradient centrifugation. Whole adult blood or buffy coats was diluted 1:2 or 1:4 respectively, in sterile Dulbecco's Phosphate Buffered Saline (PBS; Merck KGaA, German) supplemented with 2% of heat inactivated Foetal Bovine Serum (FBS; Scientifix Pty Ltd, AU) and 1mM Ethylenediaminetetraacetic acid (EDTA; Thermo Fisher Scientific Inc). 35mL of diluted blood was overlaid onto 15mL of Ficoll® Paque Plus, a density gradient medium (Cytiva, UK), in a sterile 50mL falcon tube. The tubes were centrifuged at 800xg for 20 minutes at room temperature (RT), and the centrifuge was set to decelerate with brake off to preserve the integrity of the mononuclear cell interface. The PBMC layer was recovered with a pasteur pipette and transferred to a new 50mL falcon tube. The cells were washed in sterile PBS (2% FBS & 1mM EDTA) and pelleted by centrifugation twice; first at 800xg for 5 minutes at RT and then at 400xg for 5 minutes at RT. The cell pellet was resuspended in 1ml of PBS (2% FBS & 1mM EDTA). Cell count and viability was checked using a Countess® automated cell counter (Life Technologies, Grand Island, NY, USA) by trypan blue staining (Life Technologies). Only cell preparations with viability of 90% or more were used for the subsequent experiments. Cells were resuspended at an appropriate concentration for further purification described below.

### 2.1.2 CD4<sup>+</sup> T cells isolation

CD4<sup>+</sup> T cells were isolated from buffy coats or whole blood by a negative selection rosetting strategy using RosetteSep™ Human CD4<sup>+</sup> T cell enrichment cocktail (STEMCELL Technologies). The cell isolation procedure recognizes non-CD4 lineage antigens, including CD8 (non-CD4 T cells), CD16 (natural killer cells), CD19 (B cells), CD36 (erythrocytes, platelets, monocytes -CD4 T cells) and glycophorin A (red blood cells). The cocktail was added at 1mL per 50mL of buffy or 50µL per 1mL of whole blood and incubated at RT on a rocker for 20 minutes. After incubation, CD4<sup>+</sup> T cells were isolated using density gradient centrifugation as described in Section 2.1.1 but with 1200xg instead of 800xg following the manufacturer's protocol. If cells were to be used directly for downstream analysis, only isolated CD4<sup>+</sup> T cells at a purity of greater than 90% were used. CD4<sup>+</sup> T cell purity was verified by labelling with a fluorochrome conjugated anti-CD4 monoclonal antibody (Figure 2.1) as described in Section 2.1.4.



**Figure 2.1: Purity of enriched CD4 T cells**

Example histogram plot showing >90% purity of CD4 T cells post-isolation using the RosetteSep™ Human CD4+ T cells enrichment method. Population broadcast from lymphocyte gate.

### 2.1.3 Enrichment of Treg/Tconv or Memory/Naïve populations using magnetic beads

In my PhD, two magnetic bead kits were used for further enrichment of the CD4<sup>+</sup> T cell preparation (Section 2.1.2), prior to purification of helper lineage populations, in order to reduce sorting time. For enrichment of Treg (CD4<sup>+</sup> CD25<sup>+</sup>) and T helper cells (CD4<sup>+</sup> CD25<sup>-</sup>), Human CD25 Microbeads II (Cat#130-092-983, Miltenyi Biotec GmbH, Germany) were used. For enrichment of memory CD4<sup>+</sup> (CD4<sup>+</sup> CD45RA<sup>-</sup>) and naïve CD4<sup>+</sup> (CD4<sup>+</sup> CD45RA<sup>+</sup>), Human CD45RA MicroBeads (Cat#130-045-901, Miltenyi Biotec GmbH, Germany) were used. The cells were first washed in sterile PBS wash buffer (PBS & 2% FBS & 1mM EDTA) and pelleted by centrifugation at 300xg and 10 mins. The cell pellet was resuspended in 40µL of MACS Separation Buffer (Miltenyi Biotec GmbH, Germany) per 10<sup>7</sup> cells. 10µL of CD45RA or CD25 microbeads was added and incubated for 30 minutes at 4°C. The MicroBead- conjugated cells were washed with 1–2 mL of MACS® separation buffer per 10<sup>7</sup> cells and pelleted by centrifugation at 300xg for 10 minutes and resuspended up to 1.25×10<sup>8</sup> cells in 500µL of buffer. In order to magnetically trap the targeted cells, depletion LD columns (Miltenyi Biotec GmbH, Germany) for < 10<sup>8</sup> cells were placed in a MidiMACS Separator (Miltenyi Biotec GmbH, Germany) and were rinsed with 2mL of MACS® separation buffer before applying the cell suspension. This strategy allowed for the simultaneous isolation of MicroBead-conjugated CD25<sup>+</sup> or CD45RA<sup>+</sup> (retained) cells and unconjugated CD25<sup>-</sup> or CD45RA<sup>-</sup> cells (flow through). The columns were washed 3 times with the MACS buffer to remove the non-bead bound cells from the column matrix. The CD25<sup>-</sup> or CD45RA<sup>-</sup> cells that flow through the column were collected and washed and pelleted by centrifugation at 300xg for 10 minutes at RT. Labelled CD25<sup>+</sup> or CD45RA<sup>+</sup> cells magnetically bound to the column were harvested by removal of the column from the magnetic separator and flushed out using the plunger provided in the kit, and then washed and pelleted by centrifugation at 300xg for 10 minutes at RT. The cell pellets were resuspended in CX-VIVO medium only, as indicated in Section 2.1.7 for

overnight resting and the analyses described below was carried out. Meanwhile, viability checks and cell counts were performed as described above (Section 2.1.1).

#### 2.1.4 Fluorochrome-conjugated antibody labelling of cell surface proteins

Fluorochrome-conjugated antibodies to label cell surface proteins (or markers) can be used to isolate and analyse cell subsets based on lineage and developmental stage, as well as function, using flow cytometry. Monoclonal antibodies (Table 2.1) were first titrated to determine the optimal staining concentration. Briefly,  $0.5\text{-}1 \times 10^6$  cells in  $50\mu\text{L}$  of PBS were stained per tube. Cells were incubated for 30 minutes at  $4^\circ\text{C}$ , then washed in cold PBS buffer (2% FBS) and pelleted by centrifugation at  $350 \times g$  for 5 minutes at RT. Cells were resuspended in  $0.2\text{mL}$  of the same buffer and analysed using a flow cytometer immediately.

**Table 2.1: Antibodies used for labelling surface protein for Flow Cytometry**

Specificity	Conjugate	Clone	Isotype	Manufacturer	Panel
CD4	APC-H7	SK3	Mouse IgG1, $\kappa$	BD Biosciences	Memory panel
CD25	PE-Cy <sup>TM</sup> 7	M-A251	Mouse IgG1, $\kappa$	BD Biosciences	Memory panel
CD127	PerCP-Cy <sup>TM</sup> 5.5	HIL-7R-M21	Mouse IgG1, $\kappa$	BD Biosciences	Memory panel
CD45RA	FITC	HI100	Mouse IgG2b, $\kappa$	BD Biosciences	Memory panel
CD4	BUV395	SK3	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CD25	BV421	M-A251	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CD127	PE-CF594	HIL-7R-M21	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CD45RA	APC-H7	HI100	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel

CD62L	PE	DREG-56	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel & memory panel
CXCR3	BV650	1C6	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CCR6	BV786	11A9	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CCR4	Alexa Fluor® 647	1G1	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel
CCR10	BB515	1B5	Mouse IgG1, $\kappa$	BD Biosciences	Chemokine receptor panel

### 2.1.5 Fluorochrome-conjugated antibody labelling of nuclear proteins for flow cytometry

Nuclear proteins such as FOXP3 and ZEB2, require a cell permeabilisation and fixation step to allow the fluorochrome-conjugated antibodies to enter the cell and then the nucleus of the cell. Transcription Factor Staining Buffer Set from eBioscience™ (Cat# 00-5523-00, Thermo Fisher Scientific Inc.) was used for staining nuclear proteins such as FOXP3 and ZEB2 from HEK293T/17 cells or surface-stained cells (Section 2.1.4). Prior to carrying out the staining protocol, solutions were diluted as necessary to a 1x working concentration according to the manufacturer's instructions. Cells were fixed with 1mL of Fixation/Permeabilization working solution (1:1 Fixation:Permeabilization) and pulse vortexed before incubating for 2 hours at 4°C in the dark. The cells were then washed twice in 2mL of 1X Permeabilization Buffer and pelleted by centrifugation at 500xg for 5 minutes at RT. Cells were incubated with fluorochrome-conjugated antibodies (Table 2.2) and incubated for 30 minutes at RT in the dark, then washed twice again in 2mL of 1X Permeabilization Buffer, followed by 1mL of PBS and pelleted by centrifugation at 500xg for 5 minutes at RT. Cells were resuspended in 200µL of the PBS and analysed by flow cytometry.

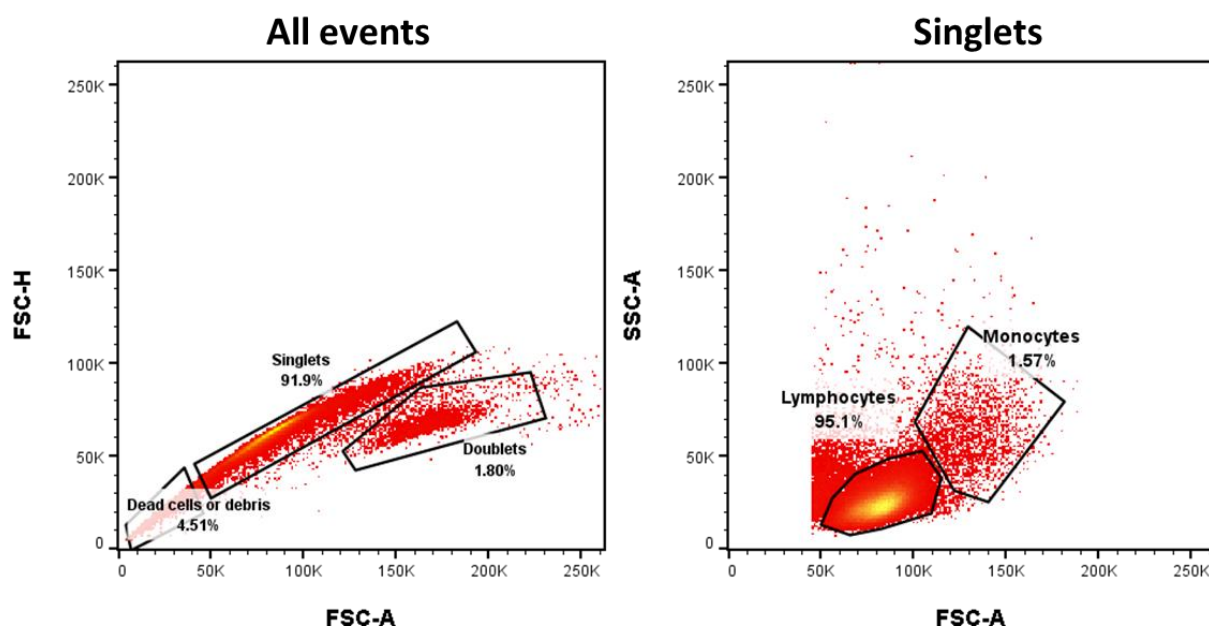
**Table 2.2: Antibodies used for labelling nuclear protein for Flow Cytometry**

Specificity	Conjugate	Clone	Isotype	Manufacturer	Location
ZEB2/SIP1	Alexa Fluor® 647	923328	Mouse IgG2b	R&D Systems	Intracellular (nuclear protein)
FOXP3	Alexa Fluor® 647	259D/C7	Mouse IgG1	BD Biosciences	Intracellular (nuclear protein)
IgG2B	Alexa Fluor® 647	133303	Mouse IgG2b	R&D Systems	Intracellular (nuclear protein)



### 2.1.6 Sort purification and analysis of T cell populations by Flow cytometry

Stained cells (Section 2.1.4 or Section 2.1.5) were analysed using a Fluorescence-activated cell sorting (FACS) Canto flow cytometer (BD Biosciences, San Jose, USA). To correct for fluorochrome generated spectral overlap, fluorescence compensation was performed using unstained cells and single-fluorochrome stained CompBeads (1:1 drop of anti-mouse Ig k and negative control) (Cat#552843, BD Biosciences, San Jose, USA) for every analysis. Further data analysis was performed using FlowJo™ 10 software (FlowJo, LLC, USA). Singlets gating was established based on forward scatter height (FSC-H) and forward scatter area (FSC-A), characteristics allowing exclusion of cell debris and doublets (Figure 2.2 left). Analysis of the singlets gate was then further refined based on side scatter (SSC-A) and FSC-A to gate on lymphocytes to exclude all monocytes (Figure 2.2 right).



**Figure 2.2: Lymphocyte gating for flow cytometry**

Flow cytometry plot showing singlets gating (left) using FSC-H and FSC-A and further gating using SSC-A and FSC-A to obtain lymphocytes. This gating strategy allowed exclusion of doublets, cell debris and monocytes from further analysis.

A FACS Fusion cytometer (BD Biosciences, San Jose, USA) was used to separate and collect cell populations of interest in sterile conditions. The gating strategy for sorting is shown in Figure 2.2. In order to maximise sort purity, a highly stringent gating strategy was used, in

which the population gates were separated from each other leaving a portion of the cells between the population gates unsorted. Cells were collected in 50% X-Vivo 15 Serum-free medium (Lonza Walkersville, Inc., USA) and 50% heat inactivated FBS (Scientifix Pty Ltd, AU). Cells were then washed, pelleted and resuspended in appropriate medium for further analysis. Cells were checked for viability and counted as described above (Section 2.1.1). Viable cells were then cultured or treated appropriately for further use. Cells sorted for RNA-based gene expression analysis, were resuspended in 700 $\mu$ L of QIAzol<sup>®</sup> Lysis Reagent (QIAGEN, Germany) and mixed by vortexing for 15 seconds to lyse the cells, and then stored at -20°C for RNA extraction.

### 2.1.7 CD4<sup>+</sup> T cell culture

Purified CD4<sup>+</sup> T cell populations (Section 2.1.6) were cultured and expanded in X-VIVO<sup>™</sup> supplemented with 2% 1M Gibco<sup>™</sup> N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES; Thermo Fisher Scientific Inc.), 1% L-Glutamine (Merck, German), 5% Human Serum (Merck, German) henceforth called Complete X-VIVO (CX-VIVO) with the addition of 100 units/mL IL-2 (proleukine; Novartis, Basel, Switzerland). Antigen independent signalling of the T cell receptor complex (CD3) and co-stimulation (CD28) are sufficient to activate T cells and this can be achieved by engagement of these complexes using activating antibodies. For activating CD4<sup>+</sup> T cells, anti-CD3/CD28 coated magnetic beads (Dynabeads<sup>®</sup> Human T-Activator CD3/CD28; Life Technologies), at a 1:1 bead-to-cell ratio were used. For resting cells after activation, beads were dissociated from cells and removed by pipetting at least 20 times and then applying a DynaMag<sup>™</sup>-2 magnet (Cat#12321D, Thermo Fisher Scientific Inc). Cells were washed twice with PBS and then these cells were transferred to a new tube. Cells were pelleted and resuspended in culture medium (CX-VIVO + 100 units/mL IL-2). All cell culture incubations were carried out at 37°C with 5% CO<sub>2</sub>.

### 2.1.8 Th1 polarisation

Naïve CD4<sup>+</sup> T cells polarise to a Th1 phenotype in the presence of IL-12. Inclusion of anti-IL4 antibody limits contamination with Th2 cells. Isolated Naïve CD4<sup>+</sup> T conv cells were activated using CD3 and CD28 antibody coated magnetic beads and cultured in medium (CX-VIVO + 100 units/mL IL-2) containing either Th1 polarisation reagents from the Human Th1 Cell Differentiation Kit (R&D Systems, US) or without, for 5 days at 37°C following the manufacturer's protocol. At day 5, cells were removed from the antibody-coated magnetic beads and expression of the Th1 defining cytokine, IFN $\gamma$ , was measured to assess the proportion of cells polarised from naïve CD4<sup>+</sup> T cells to Th1 cells (see Section 2.1.9 below for details).

### 2.1.9 Analysis of intracellular cytokines

Expression of IFN $\gamma$ , an intracellular cytokine, was analysed using flow cytometry. Cells were activated under polarising and non-polarising conditions as described in Section 2.1.8 above. Immediately prior to labelling cells were re-activated using Phorbol 12-Myristate 13-Acetate (PMA, final concentration 100ng/mL) and Ionomycin (final concentration 750ng/mL, Sigma-Aldrich) for 4hrs at 37°C. Additionally, to block intracellular transport processes, cells were incubated with GolgiStop™ (BD Biosciences, San Jose, USA) at 0.667 $\mu$ L/mL. After 4hrs, cells were washed and stained for surface antigens as described in Section 2.1.4. Before intracellular cytokine labelling, 1x10<sup>6</sup> of cells were fixed with 250 $\mu$ L of Cytofix/Cytoperm™ (BD Cytofix/Cytoperm™ Kit, Cat#554714, BD Biosciences, San Jose, USA) and mixed by vortexing before incubating for 20 minutes at 4°C to trap secretable proteins and enable antibody penetration. The cells were then washed twice in 1mL of 1X Perm/Wash™ solution (1:10 dilution of 10X stock with Milli-Q® water) and pelleted by centrifugation at 500xg for 5 minutes at RT. Cells were incubated with Alexa Fluor® 647-conjugated anti-human IFN $\gamma$  (#Cat 563495, #Clone 4S.B3, BD Biosciences, San Jose, USA) for 30 minutes at 4°C in the dark. Cells were then washed twice in 100 $\mu$ L of 1X Permeabilization Buffer and pelleted by

centrifugation at 500xg for 5 minutes at RT. Cells were resuspended in 0.2mL of the PBS and analysed immediately using the FACS Canto flow cytometer (BD Biosciences, San Jose, USA). Unstained and non-activated cells were used as controls.

## 2.2 Molecular Biology

### 2.2.1 Isolation of RNA

Cells for RNA isolation were lysed in QIAzol® Lysis Reagent (QIAGEN, Germany) (as described in Section 2.1.6) and RNA was isolated according to manufacturer's protocol using the miRNeasy Micro kit (Cat# 217084, Qiagen, Germany). Briefly, after incubation in 700µL QIAzol® Lysis Reagent (QIAGEN, Germany) for at least 5 minutes, 140µL chloroform was added and the lysates were mixed vigorously by vortexing for 15 seconds. Samples were incubated for 3 minutes at RT and then centrifuged at 12,000g for 15 minutes at 4°C. The upper aqueous layer, containing RNA was then carefully collected and transferred to a clean microfuge tube. 1.5 volumes of 100% Ethanol was added and mixed thoroughly by vortexing. 700 µ l of this was then applied to an RNeasy MiniElute spin column (included in the kit) and centrifuged at  $\geq 8,000xg$  for 15 seconds at RT. The flow-through was discarded and the process was repeated with any remaining lysate. Wash buffer concentrate RWT (included in the kit) was diluted with 30ml 100% Absolute Ethanol (analysis grade) and 700 µ L was added to the spin column and centrifuged at  $\geq 8,000xg$  for 30 seconds. The flow through was discarded. The wash buffer RPE (included in the kit) concentrate was diluted with 44ml 100% Absolute Ethanol (Analysis grade) and 500µL was added to the spin column and centrifuged as before. 500µL of freshly made 80% Ethanol (using 100% Analysis grade Absolute Ethanol and RNase-free H<sub>2</sub>O), was added and the column was centrifuged for at  $\geq 8,000xg$  for two minutes and the flow through was discarded. The membrane of the spin column was then dried by placing in a new 2 ml collection tube and centrifuged at full speed for five minutes. Finally, the

column-bound total RNA was then eluted with 14 $\mu$ L of RNase-free water, into a new collection tube, by centrifugation at full speed for one minute. The eluate was collected and used to repeat the elution step two more times to ensure maximum recovery of RNA. The resulting RNA was stored at -80°C until further use.

### 2.2.2 Nucleic acid quantification

For measuring gene expression for Quantitative Real Time PCR (qRT-PCR), the concentration of RNA was determined by UV spectrophotometry (NanoDrop™ ND-1000 spectrophotometer, Thermo Scientific, DE USA). For greater accuracy, the RNA concentration required to make RNA-seq libraries was determined using more sensitive fluorimetry (Invitrogen™ Qubit™ 3 Fluorometer, Thermo Scientific, DE USA).

For quantitation with NanoDrop™, first the background was set using 2 $\mu$ L of RNA or DNA resuspension buffer. 2 $\mu$ L of the sample was then loaded and the concentration of the sample determined. The ratios of absorbance at 260nm ( $A_{260}$ ), 230nm ( $A_{230}$ ) and 280nm ( $A_{280}$ ) are used to assess the purity of DNA and RNA. For convenience a ratio of  $\sim 1.8 A_{260}/A_{280}$  is generally accepted as “pure” for DNA; a ratio of  $\sim 2.0 A_{260}/A_{280}$  is generally accepted as “pure” for RNA. If the ratio is appreciably lower in either case, it may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280nm.

For RNA quantification with Qubit™ assay, the Qubit® RNA HS Assay Kit (detects 5–100ng) was used. All buffers and reagents used for nucleic acid quantification were supplied in the kit and steps were performed according to the manufacturer’s protocol. Qubit® RNA HS Reagent was diluted 1:200 with provided buffer to make an HS working solution. Two standards were prepared by adding 10 $\mu$ L of each Qubit® standard to 190 $\mu$ L of the HS working solution to make a total volume of 200 $\mu$ L. Next, 1-20 $\mu$ L of samples was added to 180-199 $\mu$ L of HS working solution dependant on the concentration of the samples, followed by 3 seconds of

vortexing and incubation at RT for 2 minutes before proceeding to quantitate using a Qubit® 3.0 Fluorometer (Thermo Scientific, DE USA). Prior to measuring samples, the Qubit® 3.0 Fluorometer was calibrated using two standards.

### 2.2.3 cDNA synthesis

Isolated RNA was subjected to reverse transcription to provide a complementary DNA (cDNA) template for downstream quantitative Real Time Polymerase Chain Reaction (qRT-PCR) of gene expression using QuantiTect® Reverse Transcription Kit (Qiagen, Germany). Prior to cDNA synthesis, genomic DNA was removed following manufacturer's protocol (QuantiTect Reverse Transcription Handbook, Qiagen). Briefly, 2µL of gDNA wipeout reagent was incubated with 12µL of purified RNA (up to 1 µg) for 2 minutes at 42°C followed by incubation on ice. cDNA synthesis was then carried out. A reverse-transcription reaction mix was prepared using 1µL of Quantiscript Reverse Transcriptase, 4µL of 5x Quantiscript RT Buffer and 1µL of RT Primer Mix (random hexamer primers) provided in the kit. The DNA-depleted RNA (14µL) was transferred to a fresh microfuge tube and incubated with 6µL of the reverse transcription reaction mix for 15 minutes or 30 minutes (may increase cDNA yields) at 42°C followed by followed by incubation at 95°C for 3 minutes, to terminate the reverse transcriptase reaction. The cDNA samples were store at -20°C.

### 2.2.4 Quantitative Real-Time PCR Reactions and Analysis

For quantitative real-time PCR (qRT-PCR), primer pairs were designed to detect specific mRNA derived cDNA transcripts (see

Table 2.3 below). These were designed to amplify a 50-160bp region of cDNA, usually spanning two or more exons boundaries in the gene to ensure that only mRNA derived regions are amplified. Primer pairs were matched for length (18-28bp), GC content (50–60%) and melting temp (50-65°C) (IDT, Integrated DNA Technologies, Inc., USA). Samples were run in triplicate with a 10µL reaction volume each. Briefly, the reactions contained 0.4µL of cDNA specific primer pair (10µM concentration), 4.6µL of cDNA (less than 20ng) and 5µL of 2X qPCR Master Mix (Cat# KR0389, KAPA SYBR® FAST qPCR Master Mix (2X) Kit, Roche, Switzerland). qRT-PCR was then performed on a Corbett Rotor-Gene 6000 (Corbett Life Sciences, CA USA). All reactions for real -time PCR were run in triplicate using a Qiagen Rotor-Gene Q and the means of the threshold cycles (Cts) for the triplicate samples were determined and used for subsequent quantitation (Simon, 2003). A standard curve, using a template diluted (6 dilutions) over 3 orders of magnitude and plotted against the resulting the Ct values, to determine amplification efficiency, was generated for all of the mRNAs and microRNAs and for the reference mRNA RPL13A. The standard curve method for relative quantitation was then used to determine the relative abundance of each mRNA or miRNA normalised to its reference gene or reference miRNA, respectively (Muller et al., 2002). RPL13A was chosen as the reference gene for mRNAs as its expression varied very little across the cell types and conditions used in these experiments (Wang et al., 2012). GraphPad Prism 8 (San Diego, CA) was used to perform basic statistical analysis. Alpha value was set to 0.05 to determine statistical significance in all tests, to ensure that a less than 5% chance of error is possible in accepting / rejecting the hypothesis. Statistical significance is denoted by an asterisk, in which \*p <0.05, \*\*p <0.01, \*\*\*p <0.001, and \*\*\*\*p <0.0001.

**Table 2.3: Primers used in qRT-PCR experiments**

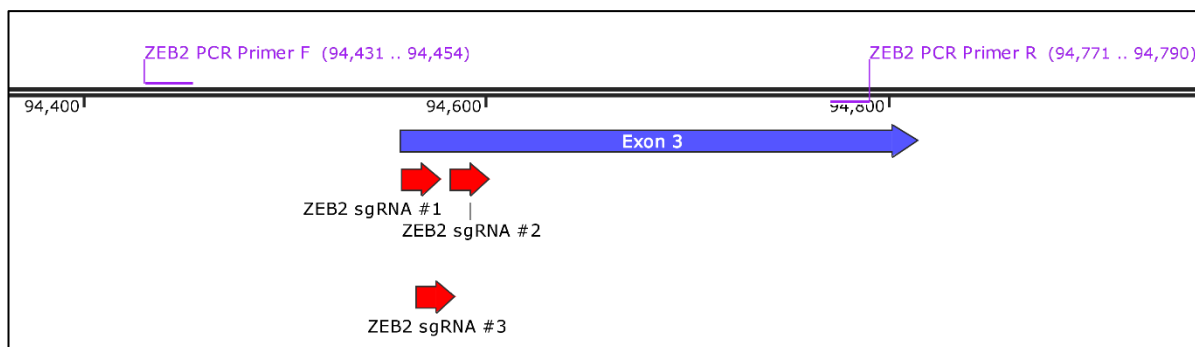
Gene	Forward (5'-3')	Reverse (5'-3')
ZEB2	AGGAGCTGTCTCGCCTTG	GGCAAAAGCATCTGGAGTTC
T-BET	AGGATTCCGGGAGAACTTTG	CCCAAGGAATTGACAGTTGG
GATA3	GCCCCTCATTAAGCCCAAG	TTGTGGTGGTCTGACAGTTCG
ROR $\gamma$ t	CTGGGCATGTCCCGAGATG	GAGGGGTCTTGACCACTGG
FOXP3	CACCACCGCCACTGGGGTCT	TCTGGGGCACAGCCGAAAGG
EpCAM	AGAACCTACTGGATCATCATTGAACTAA	CGCGTTGTGATCTCCTTCTG
S1PR5	TGCGCCTTCATCGTGCTAGAGA	TGCCAGCAGATCCGACAACGTG
HMOX1	CCATAGGCTCCTTCCTCCTTTC	GGCCTTCTTTCTAGAGAGGGAAT
IL-10	GACTTTAAGGGTTACCTGGGGTTG	TCACATGCGCCTTGATGTCTG
ROR $\gamma$ t	CTGGGCATGTCCCGAGATG	GAGGGGTCTTGACCACTGG
IFNG	TCGGTAACTGACTTGAATGTCCA	TGCCTTCCCTGTTTTAGCTGC
RPL13A	CGAGGTTGGCTGGAAGTACC	CTTCTCGGCCTGTTTCCGTAG



## 2.3 CRISPR/Cas9 Knockout

### 2.3.1 CRISPR Design

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) design tool (<https://design.synthego.com/#/>) from Synthego was used to design chemically modified synthetic guide RNA (sgRNA) targeting the ZEB2 gene in human as shown in Figure 2.3. The CRISPR design tool first identifies and prioritizes exons that are common across multiple transcripts or splice variants, enabling design of guide RNAs that target all of them. The CRISPR sequences were then narrowed down to exons in the 5' coding region of ZEB2 with a high probability of generating a complete functional knockout through insertions or deletions (INDELS). These indels in the 5' part of the gene are likely to lead to frameshift mutations that disrupt the translation of the RNA transcript, creating a full knockout of the gene/protein. All the potential off-target sites across the same genome and the number of mismatches were determined between each target CRISPR sequences and all of its potential off-target sites. Three CRISPR sequences targeting ZEB2 were then selected based on the highest likelihood of knocking out the target gene within the genome and with the least number of off-target effects (Table 2.4 & Figure 2.3). Two approaches were used to carry out ZEB2 knockout experiments. Initially one sgRNA was used (ZEB2 sgRNA #1) (in clonal expansion experiments) but this was refined for subsequent experiments where three sgRNAs were used (ZEB2 sgRNAs #1, #2 and #3), since this improves knockout efficiency and reduces experimental time. Primers were designed to increase knockout efficiency using multiple ZEB2 targeting sgRNAs. PCR primers used in genotyping to validate knockout were also synthesized by Integrated DNA Technologies.



**Figure 2.3: Schematic of 3 sgRNAs and PCR screening primers around ZEB2 exon 3**

**Table 2.4: sgRNAs sequence used in CRISPR/Cas9 ZEB2 knockout experiments**

ZEB2 sgRNA #1	GGUGAACUAUGACAAUGUAG
ZEB2 sgRNA #2	UAUGACAAUGUAGUGGACAC
ZEB2 sgRNA #3	CACAGGUUCUGAAACAGAUG

### 2.3.2 Nucleofection

CRISPR/Cas9 knockout experiments were carried out on cells sorted as described in Section 2.1.6. For efficient transfection of CRISPR-sgRNA/Cas9 complexes into the nuclei of primary human T cells, the Human T Cell Nucleofector™ Kit (Cat#VPA-1002, Lonza, Switzerland) containing the necessary buffers, was used and electroporation was carried out using the Nucleofector™2b (Cat#AAB-1001, Lonza, Switzerland) electroporation unit following the manufacturer's protocols. For knockout experiments using the single sgRNAs approach, cells were activated and expanded for 2 weeks, whereas for the triple sgRNAs approach, cells were activated for 3 days (Section 2.1.7) and divided into samples for ZEB2 targeting or control reactions. Samples nucleofected with ZEB2 targeting sgRNA(s) (Table 2.4) are knockouts (KO) whereas samples nucleofected with non-targeting control sgRNA (Negative Control sgRNA (mod) #1, Synthego, California, US) are wildtype (WT). 1.5mL of CX-VIVO culture medium (100 units/mL IL-2) per well of a 24-well plate was pre-warmed for 45 minutes in the incubator at 37°C with 5% CO<sub>2</sub> prior to the nucleofection. In a PCR tube, 6μL of 40μM

chemically modified sgRNA(s) (Synthego, California, US, described in Section 2.3.1), 4 $\mu$ L of 20 $\mu$ M 2NLS Cas9 nuclease (Synthego, California, US) and 2 $\mu$ L Human T Cell Nucleofactor™ Solution were gently mixed by pipetting and incubated at RT for at least 10 minutes to allow formation of ribonucleoprotein (RNP) complexes before returning to ice. 0.6–2 $\times$ 10<sup>6</sup> cells were then resuspended in 100 $\mu$ L of Human T Cell Nucleofactor™ Solution (Lonza, Switzerland) containing 12 $\mu$ L of RNP complex (WT or KO). The cells/RNP mix was then gently transferred to Nucleofection cuvette (Lonza, Switzerland). Cells were electroporated using a Nucleofactor™ 2b Device (Cat#AAB-1001, Lonza, Switzerland) using T-020 pulse setting for activated T cells. After nucleofection, 500 $\mu$ L of prewarmed media was gently added to the cuvette and then this was added to the remaining pre-warmed medium. Nucleofected cells were incubated at 37°C and half of the culture medium was replaced with fresh culture medium the following day. 48 hours post-nucleofection, live cells were sorted (Section 2.1.6), genomic DNA was isolated as described below, in Section 2.3.4 and then RNA was isolated as described above in Section 2.2.1.

### 2.3.3 Single Cell Clone Expansion

Optimal expansion of single cell clones was achieved using mitomycin-treated PBMC (attenuated for use as stimulatory/feeder cells) in culture medium containing anti-CD3 antibodies. PBMCs were isolated from buffy coats or fresh blood as per Section 2.1.1. Pelleted cells were then resuspended in RPMI-1640 medium (Cat#R8758, Sigma-Aldrich, MERCK, Darmstadt, Germany) + 10% FBS + 20 $\mu$ g/mL Mitomycin C (Sigma) and incubated at 37°C with 5% CO<sub>2</sub> for 1 hour for inactivation, followed by 3 washes with RPMI 1640 medium + 10% FBS. Inactivated PBMCs (1 $\times$ 10<sup>6</sup> cells/mL) were resuspended in 100 $\mu$ L of CX-VIVO containing 500 Units/mL IL-2 in a 96-well round-bottomed plate. CRISPR/Cas9 ZEB2 knockout or wild-type Th1 EM cells (described in Section 2.3.2) were labelled with APC-H7 conjugated anti-CD4 antibody (Table 2.1) and single cell sorted as per Section 2.1.5 and Section

2.1.6 into the plate containing inactivated PBMC. After sorting, a further 100 $\mu$ L of CX-VIVO with 500 Units/mL of IL-2 and 100ng/mL of anti-CD3 (functional grade OKT3 clone; eBioscience™, Thermo Fisher Scientific, US) was added. The cells were incubated at 37°C with 5% CO<sub>2</sub> for 4-8 weeks with half of the medium replaced with fresh medium containing anti-CD3 every 7-8 days.

### 2.3.4 Determination of Knockout Efficiency

Genomic DNA was extracted from at least 5x10<sup>4</sup> cells. Cells were first pelleted, supernatant aspirated and 50 $\mu$ L of QuickExtract™ DNA Extraction Solution (Cat#QE0905T, Epicentre, Lucigen, Wisconsin, US) was added and mixed by pipetting. The DNA extract was then transferred to a PCR tube and placed in a thermal cycler using the program below (Table 2.5). This program digests proteins, suspends lipids, and breaks down cellulosid material to release the DNA. Genomic PCR was then carried out (or lysates stored at -20°C for up to 1 week).

**Table 2.5: Thermal cycler program for genomic DNA extraction**

Temperature	Time
68°C	15 minutes
95°C	10 minutes
4°C	HOLD

Genomic PCR to amplify the ZEB2 region targeted by the sgRNAs was carried out using the PCR reaction mix and program below (Table 2.7 and Table 2.8). A ZEB2 PCR primer pair comprising the forward (F) and reverse (R) primers in Table 2.6 below were used to amplify the genomic region targeted by sgRNAs (Figure 2.3). The PCR amplification reaction was carried out using a T100™ Thermal Cycler (Bio-Rad Laboratories, Inc., USA)

**Table 2.6: PCR primers sequence used in CRISPR/Cas9 ZEB2 knockout experiments**

ZEB2 PCR Primer F (5'-3')	TTTCCTGACATGGTTGAGTAATTC
ZEB2 PCR Primer R (5'-3')	AATCTCGTTGTTGTGCCAGG

**Table 2.7: Preparation guide for PCR reaction**

Reagent	Volume per reaction (μL)
2X AmpliTaq Gold 360 master mix	25
Forward primer (ZEB2 PCR Primer F, 10μM)	0.5
Reverse primer (ZEB2 PCR Primer R, 10μM)	0.5
Genomic DNA	2
UltraPure water	22
Total volume	50

**Table 2.8: PCR thermocycler program**

Stage	Temperature	Time	Cycles
Enzyme activation	95°C	10 seconds	1X
Denature	95°C	30 seconds	40X
Anneal	55°C	30 seconds	
Extend	72°C	30 seconds	
Final extension	72°C	7 minutes	1X
Hold	4°C	HOLD	1X

Following amplification, the PCR product was electrophoresed using a 1% agarose gel (70V 30-45 minutes) for verification of size and amplification specificity. Amplicons were purified with NucleoSpin™ Gel and PCR Clean-up Kit (Cat#740609, Macherey-Nagel™, Düren,

Germany) as per the manufacturers' protocols. Yield and purity were measured using a NanoDrop spectrophotometer (Thermo Scientific, DE USA) as per Section 2.2.2, with samples stored at -20°C until use.

### 2.3.5 Sequencing of ZEB2 Genomic DNA Targeted by sgRNAs

Purified DNA (PD) from the genomic DNA purification reactions above (Section 2.3.4) was analysed to determine the efficiency of ZEB2 target gene knockout. DNA sequencing reactions were set up using the recommended DNA quantities in Table 2.9 below, mixed with 1µL of 10µM ZEB2 seq Primer F (Forward primer 5'-3': TTCCTGACATGGTTGAGTAATTC, 10µM) (Table 2.6) and made up to a total volume of 12µl with PCR grade H<sub>2</sub>O, total volume of 12µL. Samples were then sequenced using Sanger sequencing (Australian Genome Research Facility (AGRF), Adelaide, South Australia, Australia).

**Table 2.9: Recommended amounts of DNA and primer for sequencing reactions**

Template	Recommended Quantity for PD Samples
PCR Product 100 – 200bp	3 - 8ng
PCR Product 200 – 400bp	6 - 12ng
PCR Product 400 – 600bp	12 - 18ng
PCR Product 600 – 800bp	18 - 30ng
PCR Product >800bp	30 - 75ng
Plasmid, Single-stranded	150 - 300ng
Plasmid, Double-stranded	600 - 1500ng
Primer Quantity (one primer per reaction)	10µM

### 2.3.6 Inference of CRISPR Edits (ICE) Analysis

Genome editing efficiency was quantitatively assessed using an online tool: Inference of CRISPR Edits (ICE) (Synthego, Synthego, California, US). This software tool reports potential

editing outcomes from single and multi-guide sgRNAs (up to three sgRNAs per gene). ICE compares the Sanger sequencing traces of PCR products (Section 2.3.5) generated from genomic DNA isolated from both the knockout and Wild-Type cells. The ICE report (Figure 2.4) indicates the percentage of insertion/deletions (INDEL) and the Knockout-Score (the percentage of outcomes that lead to a putative knockout). This latter metric includes only frameshift-inducing INDELS and INDELS that are 21+ bp. Further experiments were performed on samples with more than >65% indel efficiency.



**Figure 2.4: ICE summary window**

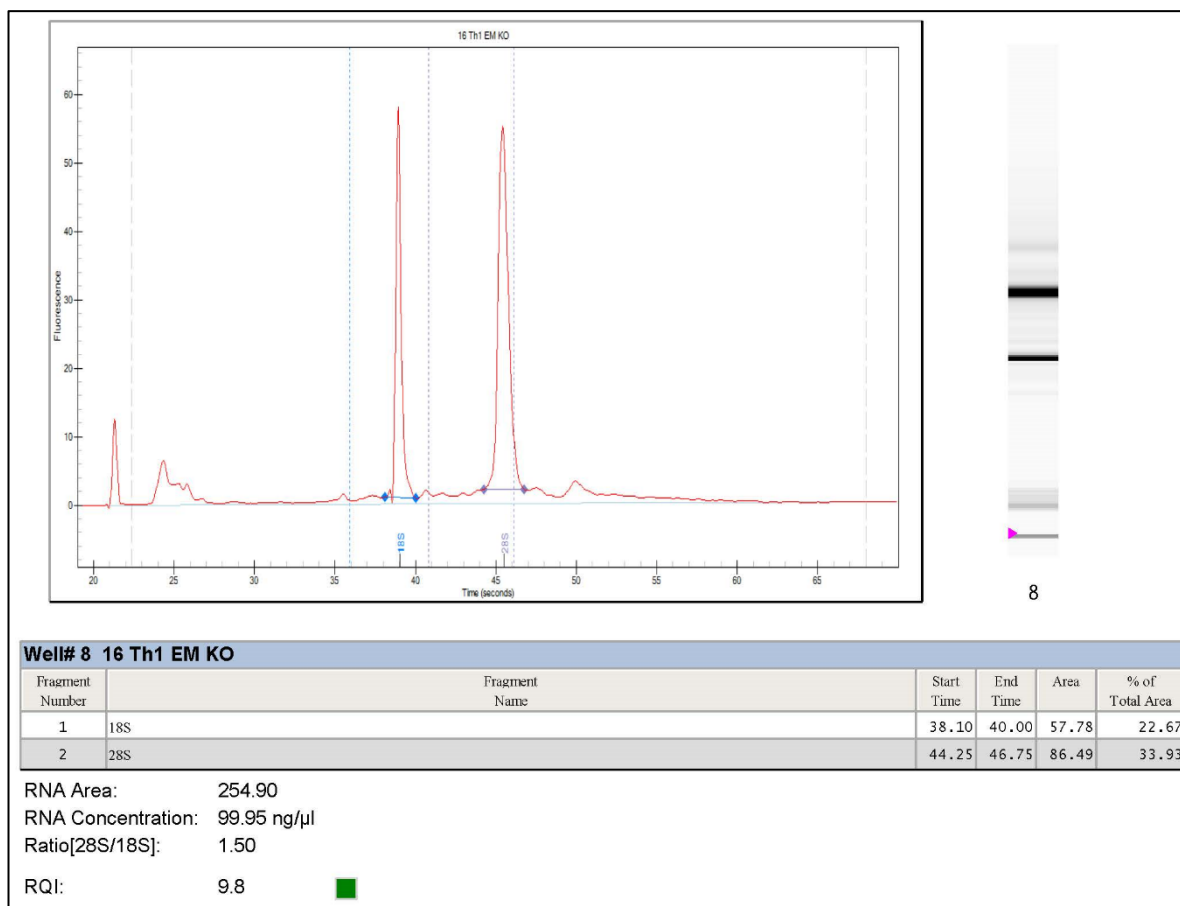
Example ICE analysis summary window of multiple samples edited by three sgRNAs targeting ZEB2. The window shows information about the edited samples, including the Indel % and Knockout-Score (% sequences that are putative knockouts).

## 2.4 RNA-Sequencing

### 2.4.1 Assessment of RNA Quality Assessment for RNA Sequencing

RNA sequencing (RNA-Seq) was used to analyse the transcriptome profile of ZEB2 knockout and Control T conv cells. RNA quality was assessed prior to generating a library for RNA seq. The quality of RNA was measured using an analysis kit (Experion™ RNA analysis kit Cat#7007103, Bio-Rad Laboratories, Inc., California, United States) and measured using an automated electrophoresis system (Experion™, Bio-Rad Laboratories, Inc., California, United States) following the manufacturer's protocol. This estimates sample integrity using gel electrophoresis and analysis of the ratios of 28S to 18S ribosomal bands, producing an index of relative quality (an RNA Integrity Number (RIN) score between 1 and 10, with 10 being the highest quality samples showing the least degradation). RNAs with RIN score >7 were used for RNA-seq library generation below.





**Figure 2.5: Electropherogram of an RNA sample used for RNA-seq**

An example electropherogram to determine RNA integrity number (RIN) of samples using the Experion™ Automated Electrophoresis System.

#### 2.4.2 Selection of Poly(A) mRNA

Poly-A-tailed messenger RNA (mRNA) selection was carried out to allow efficient detection of mRNAs. The NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs®, Massachusetts, United States) was used to select for intact mRNA from previously isolated total RNA (Section 2.2.1) with a RIN score of >7 (Section 2.4.1), following the manufacturer's protocol. The amount of total RNA ranged from 10ng–1μg which was quantified by Qubit Fluorometer as described in Section 2.2.2.

#### 2.4.3 RNA-seq library generation

There are multiple steps involved in generating RNA-seq libraries. The NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina® (Cat#E7760, New England Biolabs®,

Massachusetts, United States) was used for strand specific cDNA synthesis, NEBNext® Multiplex Oligos for Illumina (New England Biolabs®, Massachusetts, United States) was used for barcoding individual RNA libraries, and SPRIselect (Cat#B23317, Beckman Coulter, California, United States) was used for size selection and purification. All buffers and reagents used in library generation were provided in the kits and manufacturer's protocols (<https://www.nebiolabs.com.au/protocols/2014/12/02/protocol-for-use-with-nebnext-poly-a-mrna-magnetic-isolation-module-neb-e74901>) were followed.

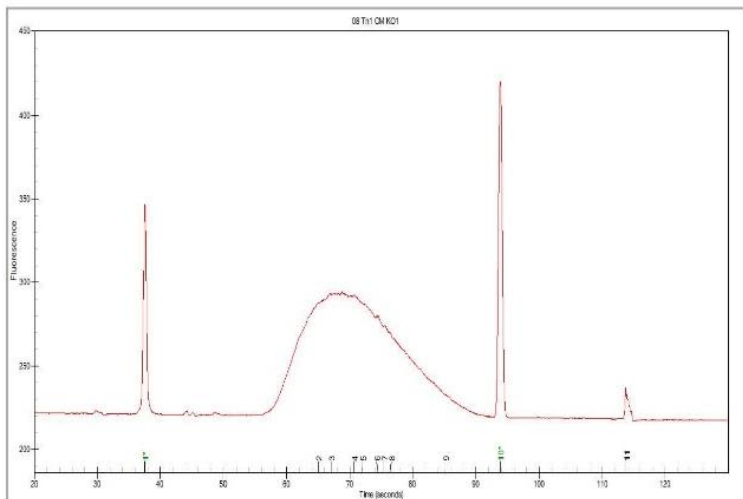
In brief, the magnetically selected polyA mRNA (Section 2.4.2) were fragmented to ~200bp of RNA insert size for 15 minutes, followed by cDNA synthesis, end repair, and adaptor ligation. After the recommended cycles of PCR, based on initial input RNA amount, these RNA-seq libraries were then analysed using the Experion™ Automated Electrophoresis System (Bio-Rad Laboratories, Inc., California, United States), using Experion™ DNA 1K Reagents Kit (Cat#7007107, Bio-Rad Laboratories, Inc., California, United States) and following manufacturer's protocol. RNA-seq libraries which showed a narrow distribution of a peak size approximately 300bp are considered good quality as indicated in Figure 2.6. Libraries were then quantified by qRT-PCR using KAPA Library Quantification Kits (Cat#07960140001, Roche, Basel, Switzerland) and pooled at equimolar concentrations. Pooled libraries that passed quality control then underwent 150bp pair-end sequencing (GENEWIZ, China) using a high-throughput sequencing system HiSeq 2500 (Illumina, Inc., California, United States). Output raw data was then demultiplexed into individual sample libraries by Illumina bcl2fastq 2.17 software based on individual libraries' barcode index information into the number of reads and quality score (Q30).

Egram, Gel Lane and Result Table Report

Page 7 of 25

**Project:** Vincent RNA library  
**Assay:** DNA 1K  
**Run:** Th1 CM EM WT KO Pool Run 1\_6-12-2019\_1-36-48 PM  
**Run Version:** N/A  
**Acq. Analyst:** DefaultUser  
**Acq. Time:** 6/12/2019 1:36:49 PM  
**Signature:** N/A

Well# 2 08 Th1 CM KO1



2

Well# 2 08 Th1 CM KO1

Peak State	Peak Number	Mig. Time (secs)	Size (bp)	Corrected Area	Area Ratio	Concentration (ng/ul)	Molarity (nmole/l)	% Total	Observation	Comments
	1	37.55	15	193.78					Lower Marker	
	2	65.03	300	436.45	2.9909	9.73	49.07	33.15		
	3	66.99	326	103.41	0.7086	2.29	10.66	7.81		
	4	70.74	374	122.11	0.8368	2.68	10.84	9.12		
	5	72.03	391	185.34	1.2701	4.05	15.70	13.79		
	6	74.50	432	62.13	0.4258	1.33	4.66	4.52		
	7	75.48	449	67.45	0.4622	1.42	4.80	4.85		
	8	76.35	465	334.96	2.2954	7.00	22.82	23.86		
	9	85.21	684	39.33	0.2696	0.85	1.88	2.90		
	10	93.80	1,500	145.93					Upper Marker	
	11	113.77	0	9.17						

**Sample Peaks Found:** 8  
**Total Concentration:** 29.4 ng/ul  
**Total Molarity:** 120.4 nmol/l

**Figure 2.6: RNA-seq library size distribution on an Experion™ Automated Electrophoresis System**

Example electropherogram shows a narrow distribution with a peak size approximately 300bp

#### 2.4.4 Quality Control Assessment

RNA-seq raw data in FASTQ format from individual libraries were uploaded to Phoenix High Performance Compute Server, hosted by The University of Adelaide for processing. FastQC is a common tool used for checking the quality of the FASTQ file (Andrews, 2010). A FastQC report was generated in html format for visualization and aggregated using an R package called ngsReports. After inspection of data quality, low quality bases and adapters were then removed using AdapterRemoval (Lindgreen, 2012).

#### 2.4.5 Pseudoalignment

Quantification of transcript expression in RNA-seq data was conducted using Salmon version 1.2.1 (Patro et al., 2017). First, a transcriptome index was built using transcript sequences and primary assembly genome sequences from Human Release 34 (Frankish et al., 2019). The index constructed once, can then be reused to quantify many experiments. RNA-seq reads (Section 2.4.4) were then pseudoaligned to the transcriptome index and output was generated in a sub-directory containing a simple TSV format file listing the name of each transcript, its length, effective length, and its abundance in terms of Transcripts Per Million (TPM) and estimated number of reads originating from this transcript. After the quantification, results can be imported into R statistical software environment (R Core Team, 2020) for downstream analysis.

#### 2.4.6 Differential Gene Expression Analysis

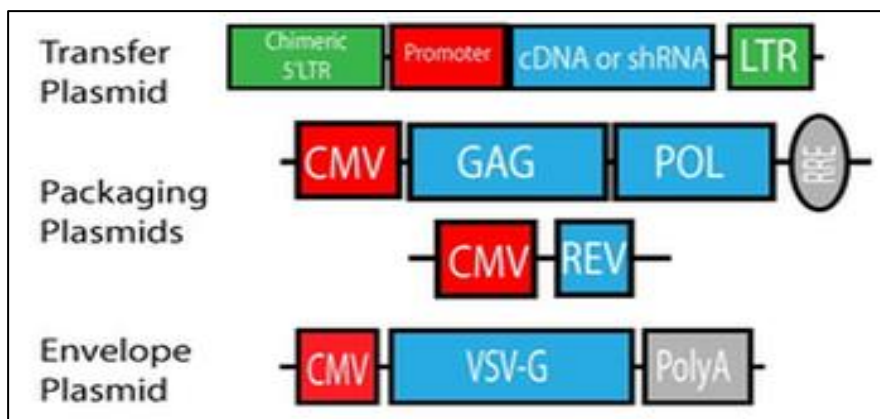
Differential expression (DE) was carried out using edgeR (Robinson et al., 2010) and RUVseq (Risso et al., 2014) to compare the transcriptome between groups. Transcript-level counts were summarized into gene-level counts and form a DGEList object containing raw counts, sample information and gene information. Genes with more than 1 count per million (CPM) in minimum samples of an experimental group were retained. Sample library sizes were normalized with calcNormFactors function. A set of negative control genes were obtained by taking genes that were least significantly DE, based on first-pass DE analysis performed with

the negative binomial GLM approach. Factors of unwanted variation was then calculated with the RUVg function and added into a design matrix with experimental grouping information, prior to performing a final DE analysis. Genewise statistical tests for contrast between two groups relative to a log<sub>2</sub>-fold-change (log<sub>2</sub>FC) threshold of 1.1 and false discovery rate (FDR) of <0.05 were conducted using glmTreat function. DE genes were extracted for visualization, pathway enrichment analysis, data intersection or modular co-expression analysis.

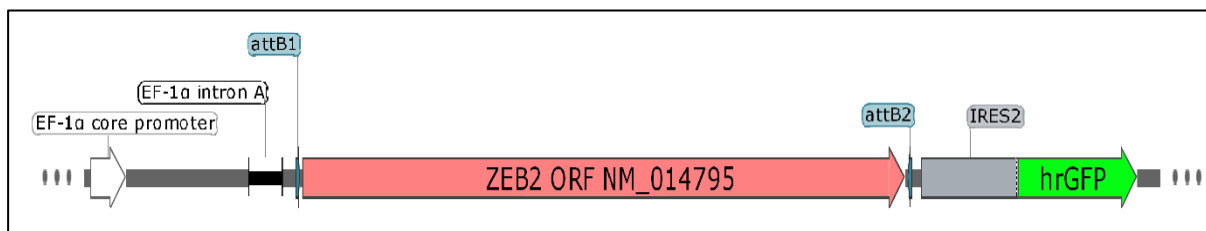
## 2.5 Overexpression of ZEB2 with lentivirus

### 2.5.1 Plasmids for packaging HEK293T/17 cells with LV411 lentivirus

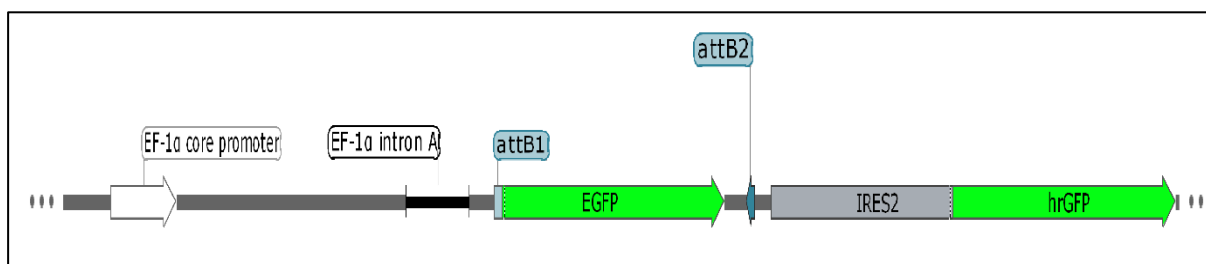
A third-generation lentiviral (LV) system was used to express exogenous genes in human T cells. There were 4 plasmids in the system consisting of the transfer plasmid containing the transgene flanked by the viral LTRs (LV-411-ZEB2-IRES-hrGFP or LV-411-GFP-IRES-hrGFP), the envelope plasmid, in this case the broadly tropic VSV-G (p-CMV-VSV-g), the GAG/POL (ps-PAX2) plasmid encoding the viral core proteins (GAG) and the reverse transcriptase enzymes (POL) and the REV plasmid (p-RSV-REV) essential for regulating expression of virus proteins (Figure 2.7). The LV transfer plasmid construct described previously (Brown et al., 2010, Brown et al., 2018, Barry et al., 2001) encodes full length ZEB2 (Figure 2.8) or enhance green fluorescent protein (eGFP) (Figure 2.9 transcribed from the EF1 $\alpha$  promoter. An Internal Ribosome Entry Site (IRES) allows translation of a second gene from the same promoter to act as a marker; humanised Renilla reniformis green fluorescent protein (hrGFP).



**Figure 2.7: Components of third-generation lentiviral plasmids**  
Third Generation Lentiviral Plasmids. Figure adapted from Boris Fehse Lab Plasmids – Addgene.



**Figure 2.8: LV411-ZEB2-IRES-hrGFP transfer plasmid construct**



**Figure 2.9: LV411-eGFP-IRES-hrGFP transfer plasmid construct**

### 2.5.2 Plasmid Purification

Plasmid DNA for cloning (NucleoBond™ Xtra Midi EF kit (Macherey-Nagel™, Germany) and sequencing (NucleoSpin® Plasmid Mini kit (Macherey-Nagel™, Germany) was purified from bacterial cultures following the Manufacturer's protocols. DNA yield and purity was quantified using a NanoDrop™ ND-1000 spectrophotometer (Thermo Scientific, DE USA) as described in Section 2.2.2, and stored at -20°C until use.

### 2.5.3 Thawing of HEK293T/17 Cells

Early passage (16-25) HEK293T/17 cells were rapidly thawed by placing the cryovial in a 37°C water bath and swirling gently. The cryovial was then removed from the water bath whilst ice crystals remained. The cryovial was then sterilized by spraying with 70% ethanol. A 1-mL or 5-mL pipette was used to transfer thawed cells drop-wise into 9 mL pre-warmed Dulbecco's Modified Eagle's Medium - high glucose (DMEM; Cat#5671, Sigma-Aldrich, MERCK, Darmstadt, Germany) with 10% FBS in a 15-mL centrifuge tube and gently mix by pipetting up and down several times. Cells were pelleted at 170 xg for 5 minutes and resuspended with fresh pre-warmed DMEM (10% FBS). Resuspended cells were seeded into Corning® T-75 flasks (Cat#430641) and incubated at 37°C with 5% CO<sub>2</sub>. Cells were checked for viability and counted as described above (Section 2.1.1).

### 2.5.4 HEK293T/17 Cell Culture and Preparation

Thawed HEK293T/17 cells, obtained from ATCC (Manassas, Virginia, USA) (Section 2.5.3) were maintained in DMEM (10% FBS) for 2-4 weeks prior to packaging. Cells were constantly monitored for medium exhaustion and subcultured every 2-3 days at a recommended ratio of 1:4 to 1:8 to prevent overgrowth (maximum 80% confluency) until cells were ready for packaging. For subculturing of HEK293T/17 cells, culture medium was removed and discarded. Briefly, the cell layer was washed with PBS followed by incubation at 37°C with 1 mL of 1x trypsin solution (Cat#T4549, Sigma-Aldrich) for 2 minutes. The cell layer was then dispersed with a sharp tap and the trypsin inactivated using approximately 10 x trypsin volume of DMEM (10% FBS) and gently mixed. Cells were then pelleted at 170xg for 5 minutes and resuspended with fresh pre-warmed DMEM (10% FBS). Appropriate aliquots of the cell suspension were then seeded into a new flask.

### 2.5.5 Packaging and Concentration of Lentiviral Supernatant

One day prior to transfection,  $9 \times 10^6$  HEK 293T cells were seeded into a Corning® T-75 flasks (Cat#430641) in 16 mL of LV packaging medium (Opti-MEM™, (Cat#51985034, Gibco™, Thermo Fisher Scientific Inc.) + 5% FBS (Scientifix Pty Ltd, AU) + 1mM Sodium pyruvate (Cat#11360070, Gibco™, Thermo Fisher Scientific Inc.). Immediately prior to transfection, 8ml of the LV packaging medium was removed and the cells were transfected with the 4 plasmids as described above, using LV transfer plasmid (6.5µg), VSV-G plasmid (3.75µg), Gag/Pol plasmid (7.5µg) and REV plasmid (6.25µg) with Lipofectamine™ 3000 transfection reagents (Cat#L3000015, Invitrogen, CA USA) according to the manufacturer's protocol. Briefly, two reaction mixes were prepared separately in 10ml sterile plastic tubes for each transfection reaction, using volumes in accordance with the manufacturer's recommendations for a T75 tissue culture flask. 55µL of Lipofectamine™ 3000 was mixed with Opti-MEM™ medium to a final volume of 2ml and in a separate tube, plasmid DNAs (LV transfer plasmid, VSV-G, GAG/POL & REV) were mixed with 47µL of p3000 Enhancer Reagent and made up to 2mL with Opti-MEM™ medium. To prevent precipitation, it is important to add the DNA prior to the Enhancer Reagent. Both reaction mixes were then incubated at RT for 5 minutes. The Lipofectamine™ 3000/Opti-MEM™ mix was gently mixed with Enhancer Reagent /DNA/Opti-MEM™ mix and incubated at RT for 20 minutes before adding dropwise to cells. LV packaging medium was replaced with fresh medium 4 hours post-transfection.

Lentiviruses are sensitive to extreme temperature shifts resulting in decreased lentiviral titre. Therefore, all of the following lentivirus harvesting and handling was carried out on wet ice or 4°C. All of these procedures are carried out using PC2 protocols and to standard PC2 laboratory practice, GMO (3rd generation self-inactivating lentivirus) handling was carried out. This involves wearing double protective gloves throughout the procedures, removing the outer pair of gloves only for final clean up, decontaminating all pipettes, tips and other consumables in



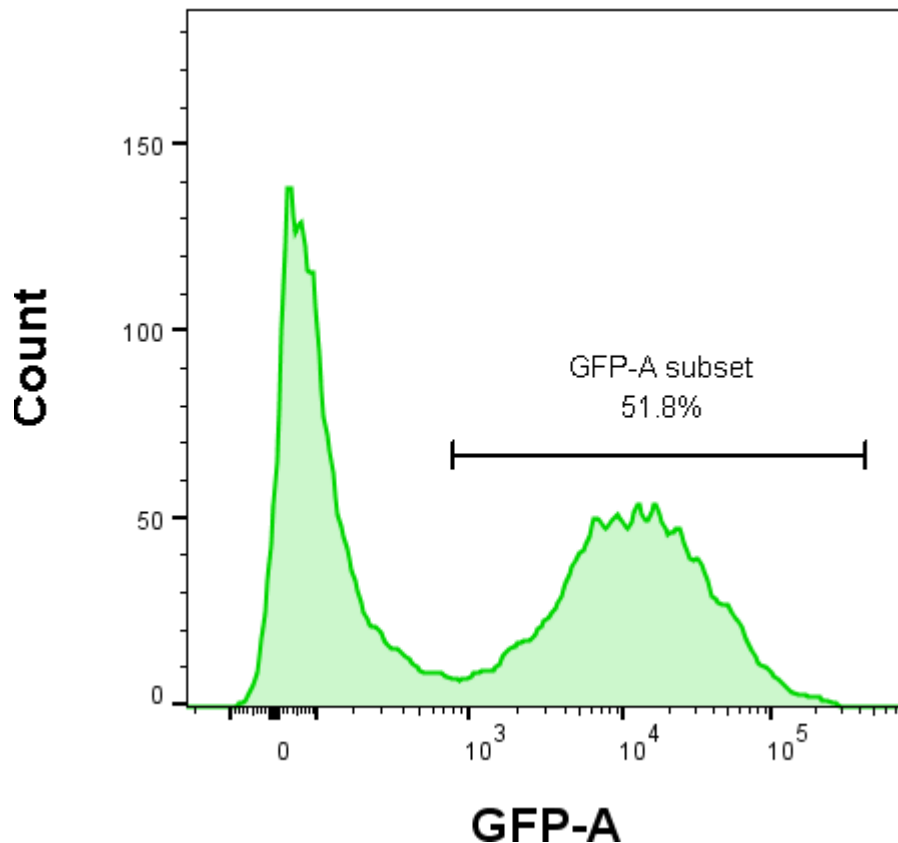
10% Sodium Hypochlorite or professional sanitizing solution LYSOL (Alkyl dimethyl benzyl ammonium chloride), prior to disposal using double waste bags intended for incineration. All surfaces are decontaminated using one of the above sanitizers followed by standard clean up with 70% ethanol.

Virus supernatant was harvested from flasks at 24 and 48 hours post transfection. 24 hours after transfection, the culture medium, containing viable virus, was very carefully removed from the flask using a plastic pipette. Great care should be taken to avoid dislodging or damaging the cell monolayer. The culture medium (virus supernatant) was dispensed into a 50ml Falcon tube and fresh, warm LV packaging medium (16mL) was very carefully added down the side wall of the tissue culture flask, to avoid dislodging the cell monolayer. The virus supernatant removed from the flask was then centrifuged at 1,200rpm for 5 minutes to pellet unwanted loose cells. All further manipulations were carried out on ice to preserve the integrity and viability of virus. Virus supernatant was then passed through a 0.45 $\mu$ M filter (remove cellular debris) into a round bottomed ultracentrifuge tube (Ultra Clear Cat#344058, Beckman Coulter, CA USA) before concentration by centrifugation (49,200xg (r av), 4°C, 90 minutes) in a swing-out rotor (SW 32 TI Rotor, Beckman Coulter, CA USA) using an Optima Ultracentrifuge (Beckman Coulter, CA USA). On ice, the virus was collected by aspirating the supernatant from the centrifuge tube, overturning the centrifuge tube and wiping the rim with a tissue to collect any remaining supernatant and finally resuspending the virus pellet in the remaining supernatant alone or with added cold culture medium (Opti-MEM™ + 5% FBS) to a final volume of approximately 100 $\mu$ L. Virus resuspension was achieved by repeatedly pipetting up and down thirty times on ice over a 15 minutes time period. Virus was then aliquoted into small volumes and both concentrated and unconcentrated virus were stored at -80°C prior to use or titre, as described in Section 2.5.6 below. Viruses are only transported in an unbreakable sealed container located in another unbreakable sealed container (double-contained).

### 2.5.6 Viral Titre

The virus titre was determined by calculating GFP+ cells in serially diluted virus-transduced HEK293T/17 cells, using flow cytometry.  $1 \times 10^5$  HEK293T cells per well (24 well plate) were transduced with (50 $\mu$ L) of serially diluted lentiviral supernatant (1:10, 1:50 and 1:100) or concentrated virus (1:100, 1:500 and 1:1000) in 500 $\mu$ L of DMEM + 10% FBS medium containing 8 $\mu$ g of polybrene (Sigma-Aldrich). All dilutions were carried out in duplicate. The following day, medium was replaced with fresh medium. Cells were harvested 48 hours post transduction by trypsinisation and transferred to a Falcon® FACS tube (Cat# 352008). Cells were then pelleted at 170xg for 5 minutes, resuspended in 100 $\mu$ L of PBS and flow cytometric analysis (Section 2.1.6) was then performed. Percentage of GFP-positive cells were determined with untransduced cells as control to set the gate (Figure 2.10). Lentiviral titre was calculated using the formula:

$$\textit{Titre} = \frac{\%GFP \textit{ positive cells} \times 1 \times 10^5 \times \textit{dilution factor}}{\textit{Volume of medium}} = \textit{virus particles/mL}$$



**Figure 2.10: Proportion of GFP+ HEK 293T cells transduced with lentivirus**

Example histogram plot showing proportion of HEK 293T cells expressing the reporter GFP. Cells were transduced with LV411-eGFP-IRES-hrGFP concentrate serially diluted at 1:500.

**2.5.7 Transduction of CD4+ T Cells with LV411-ZEB2-IRES-hrGFP**

Purified cells (Section 2.1.6) were pelleted and resuspended in CX-VIVO (100 units/mL IL-2 & 8µg/mL polybrene) and activated using anti-CD3/CD28 coated magnetic bead (Dynabeads® Human T-Activator CD3/CD28; Life Technologies), at a 3:1 bead-to-cell ratio for 1 hour before transduction. After activation, cells were transduced by adding lentivirus (Section 2.5.6) at a Multiplicity Of Infection (MOI) of x10-20 to cells and then incubated at 37°C with 5% CO<sub>2</sub> overnight. The next day, half of the medium was replaced with fresh medium without polybrene. Transduction efficiency was determined 24 hours and 48 hours post transduction by flow cytometric analysis of % GFP+ cells, as described above (Section 2.1.6 & Section 2.5.6). Transduced cells can be further be expanded to achieve cell numbers of interest (Section 2.1.7) before performing any downstream analyses.

## **CHAPTER 3: ZEB2 EXPRESSION IN CD4+ T CELLS**

### 3.1 Introduction

The effector activities of CD4<sup>+</sup> T cells are balanced by suppressive actions of Treg (CD4<sup>+</sup>, CD25<sup>hi</sup>, CD127<sup>lo</sup>) cells and together they mediate constant surveillance of the immune milieu, reacting appropriately to immune challenge. In order to achieve such finely tuned and appropriate responses, CD4<sup>+</sup> T cells are not only divided into effector and suppressive cells but into a growing, and still largely under-characterised group of lineages and subsets which allow for the highly regulated and specific array of immune responses. In the steady state, in T effector and Treg cell compartments, both naïve CD4<sup>+</sup> T cells and antigen-experienced memory cells circulate to allow quick responses to de novo or repeat immune challenge. The antigen-experienced cells are broadly divided into Central Memory (CM) and Effector Memory (EM) populations that carry out immune surveillance, circulating between the lymphoid organs or transiting to peripheral blood and tissues, respectively, poised to rapidly react to an immune assault. A much smaller population, effector memory RA<sup>+</sup> (TEMRA) cells can also be detected which share some properties associated with both Memory and Naïve T cells, although there is still debate as to whether they are an exhausted population of CD4<sup>+</sup> cells or whether they are a distinct and functional population in their own right (Tian et al., 2017).

The functional properties of the CD4<sup>+</sup> T memory compartment are defined by a transcriptional programme which enables the cell to express pathogen-specific effector molecules and homing receptors, leading to their classification as Tconv (or T helper) (Th) subsets. These subsets have been broadly defined based on expression of a “master” transcription factor that orchestrates the transcriptional programme specific for a particular immune challenge. Thus, CD4<sup>+</sup>T cells require a range of mechanisms to ensure lineage fidelity. However, the function of these cell populations: Th1, Th2, Th17 (O'Shea and Paul, 2010, Zhu and Paul, 2010), Th1/17 (Leung et al., 2010), Th22 (Plank et al., 2017), Th9 (Dardalhon et al., 2008, Veldhoen et al., 2008), Treg (Sakaguchi, 2000), and follicular helper (Tfh) (Schaerli et al., 2000) T cell subsets are not

entirely fixed in a given lineage, but somewhat plastic. Tregs express the master transcription factor FOXP3 and are critical in restraining self-reactivity and excessive inflammation. Although Treg cells are generally considered to be a separate lineage of CD4<sup>+</sup> T cells, studies in mice and human have indicated that they are able to respond to the same cues and home to the same sites of inflammation as the Tconv subsets (Duhon et al., 2012, Höllbacher et al., 2020). This suggests that, like Th cells subsets, Treg cells differentiate into specialized subsets during different types of immune responses, and that this is critical for the appropriate regulation of different Th cell populations. Whilst plasticity in T cell response is important for tailoring and attenuating an immune response, it also has important implications for autoimmune disease. Developing a highly adaptable inflammatory response at the expense of maintaining lineage fidelity can lead to autoimmune disease, as observed in mice and humans where imbalance in Th1, Th17 and Th1/17 cells contributes to IBD.

Analysis of mRNA microarray data of Treg and Tconv intersected with FOXP3 ChIP-seq in Treg, led to the identification of transcriptional factors (TFs), including Zinc finger E box binding homeobox 2 (ZEB2, SIP1, ZFXH1B) (Sadlon et al., 2010) that were differentially expressed and molecularly controlled by FOXP3. ZEB2 is a TF best known for its role in epithelial to mesenchymal transition (EMT), in which epithelial cells lose their cellular identity and are converted into mesenchymal cells (Brabletz and Brabletz, 2010). EMT transitions are crucial in embryonic development, wound healing, and cancer (Craene and Berx, 2013). Mice in which ZEB2 is constitutively knocked out are embryonic lethal (Higashi et al., 2002), while patients with heterozygous abnormalities in ZEB2 can develop Hirschsprung's disease of the large intestine and Mowat-Wilson syndrome (Vandewalle et al., 2009).

In the immune system, it has recently been reported that ZEB2 functions to regulate NK cell maturation (van Helden et al., 2015), the terminal differentiation of CD8<sup>+</sup> effector T cells (Dominguez et al., 2015, Omilusik et al., 2015), the differentiation and development of pDCs

and cDC2s (Scott et al., 2016, Wu et al., 2016) and the control of granulocyte-macrophage progenitor (GMP) fate as well as maintaining the tissue identities of macrophages (Scott et al., 2018). Additionally, ZEB2 has been suggested to play an effector role in cytotoxic CD4 (CTL) (Patil et al., 2018) whilst its expression in other CD4+ T cell subsets remains undetermined.

Characterising the distribution of ZEB2 expression across the CD4+ T cell subset population is the first step in understanding its role. Subsets that express ZEB2 may have a role in subset-associated specific function. This chapter describes the methodology used for a comprehensive characterisation of ZEB2 in CD4 subsets using anti-ZEB2 antibody or qRT-PCR of purified CD4 subsets. Examination of ZEB2 expression in a variety of CD4 subsets revealed high expression of ZEB2 Tconv EM, Tconv Th1 and specifically the Tconv Th1 EM.

## 3.2 Aims and Hypothesis

The hypothesis for this chapter is that ZEB2 is lineage restricted and predominantly expressed in effector memory CD4<sup>+</sup> T cells.

The focus of this chapter is to address Aim 1 of my PhD proposal.

- 1.1. To determine accessibility differences on ZEB2 gene locus between Treg and Tconv.
- 1.2. To determine ZEB2 expression across naïve and memory populations in Treg and Tconv.
- 1.3. To determine ZEB2 expression across helper lineages in Treg and Tconv.
- 1.4. To determine ZEB2 expression across memory populations within the helper lineages of Tconv.



### 3.3 Material & Methods

#### 3.3.1 ZEB2 Gene Locus Metagenomics

The WashU Epigenome Browser (<https://epigenomegateway.wustl.edu/>) provides visualization of the ZEB2 gene body, with integration and analysis tools for epigenomic datasets either generated by the Barry lab or in the public domain. ATAC-seq data tracks from Tconv and Treg (Wong et al., unpublished) were displayed with differential accessibility (DA) peaks. In addition, CHIP data tracks for the transcription factors FOXP3 (Sadlon et al., 2010) and T-bet (Hertweck et al., 2016) were also included to display putative binding sites in the ZEB2 gene body. In order to capture the combinatorial interactions between different chromatin marks of cell types, chromatin states of T helper, Th17 and Treg (NIH Roadmap Epigenomics, 2020) were loaded together with T cell enhancer tracks by Vahedi et al. (2015). Enumeration of these tracks allows the unbiased determination of various T cells key regulatory nodes in ZEB2 gene body.

#### 3.3.2 Detection of ZEB2 by flow cytometry in ZEB2 overexpressing HEK293T/17 cells

HEK293T/17 is a primary embryonic human kidney cell line that is transduced easily with LV411-ZEB2-IRES-hrGFP virus. These ZEB2 overexpressed (LV411-ZEB2-IRES-hrGFP) HEK293T/17 cells were generated from methods as indicated in Section 2.5.5 & Section 2.5.6. To confirm the proviral transcripts were being made after lentiviral transduction, RNA was isolated from LV411-ZEB2 and LV411-GFP (control) HEK293T/17 cells and cDNA synthesis was performed. ZEB2 mRNA expression of LV411-ZEB2 HEK293T/17 cells were compared with untransduced by qRT-PCR using ZEB2 primer pairs (sequences listed in

Table 2.3) following method in Section 2.2.4.

ZEB2 antibody labelling was first optimised and tested in LV411-ZEB2 and LV411-GFP HEK293T/17 cells. If ZEB2 protein in the ZEB2 overexpressing HEK293T/17 cells could be detected using an antibody specific for ZEB2, then it could potentially be used for detection of ZEB2 protein in the heterogeneous CD4<sup>+</sup> T cell populations. Nuclear protein labelling of ZEB2 was carried out following the eBioscience manufacturer's protocol as indicated in Section 2.1.5. Briefly,  $1 \times 10^6$  of cells were dispensed into a Falcon® FACS tube (Cat# 352008), and freshly prepared eBioscience Foxp3 Fixation/Permeabilisation buffer (1mL) was then added to each sample and mixed by vortexing. Samples were incubated at 4 °C for 30 minutes in the dark, and then washed twice with 2mL of eBioscience 1x Permeabilisation buffer. Cells were then incubated with the addition of 5µL of anti-human ZEB2 conjugated with Alexa Fluor® 647 antibody (Cat#IC73782R, clone: 923328, R&D Systems, USA) for 30 minutes at 4°C in the dark. Samples were washed twice with 2mL of eBioscience 1x Permeabilisation buffer and once with 2mL PBS. Cells were then pelleted and resuspended in 200µL PBS and stored at 4°C in the dark or analysed by flow cytometry (BD FACS Canto, BD Biosciences, San Jose, USA), and raw FCS data files were then analysed using FlowJo™ 10 software (FlowJo, LLC, USA).

### 3.3.3 Flow Cytometry Panels for Isolating CD4<sup>+</sup> T Cell Subsets

Four antibody labelling panels were designed to sort and immunophenotype different CD4<sup>+</sup> T cell subsets from the following categories: **panel 1**: Treg and Tconv panel; **panel 2**: Naïve & Memory Treg and Tconv panel; **panel 3**: Chemokine receptor panel; **panel 4**: Memory chemokine receptor panel. These 4 panels were designed to cover the majority of the CD4<sup>+</sup> T cells subsets. CD4<sup>+</sup> T cell were first enriched from buffy coats (Australian Red Cross) using RosetteSep™ Human CD4<sup>+</sup> T cells enrichment cocktail (STEMCELL Technologies) as indicated in Section 2.1.2 before labelled with different antibody panels following methods in Section 2.1.4.

For panel 1, CD4, CD25 and CD127 were used to classify the Treg and Tconv (Table 3.1). Enriched CD4<sup>+</sup> T cells were then labelled with the following antibodies: APC-H7 anti-human CD4, PE-Cy<sup>TM</sup>7 anti-human CD25 and PerCP-Cy<sup>TM</sup>5.5 anti-human CD127 (Table 2.1).

**Table 3.1: Panel 1: Treg and Tconv panel**

Cell type	Surface markers / Gating strategy
Treg	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup>
Tconv	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup>

Durek et al. (2016) reported that ZEB2 was expressed in an increasing gradient from Naive, Central Memory (CM), Effector Memory (EM) and Effector Memory RA<sup>+</sup> (EMRA<sup>+</sup>) in human CD4<sup>+</sup> T cell pools. However, the ZEB2 expression in Naïve and Memory Treg and Tconv were not assessed. Therefore, in panel 2 CD62L and CD45RA were used together with the Treg and Tconv panel above, to classify both Treg and Tconv across four maturation stages: Naive, CM, effector EM and EMRA<sup>+</sup> (Table 3.2). Enriched CD4<sup>+</sup> T cells were labelled with the following antibodies: APC-H7 anti-human CD4, PE-Cy<sup>TM</sup>7 anti-human CD25, PerCP-Cy<sup>TM</sup>5.5 anti-human CD127, FITC anti-human CD45RA and PE anti-human CD62L (Table 2.1).

**Table 3.2: Panel 2: Naïve & memory panel**

Cell type	Surface markers / Gating strategy
Tconv naive	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>+</sup> , CD62L <sup>+</sup>
Tconv CM	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CD62L <sup>+</sup>
Tconv EM	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CD62L <sup>-</sup>
Tconv EMRA <sup>+</sup>	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>+</sup> , CD62L <sup>-</sup>
Treg naive	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>+</sup> , CD62L <sup>+</sup>
Treg CM	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CD62L <sup>+</sup>
Treg EM	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CD62L <sup>-</sup>
Treg EMRA <sup>+</sup>	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>+</sup> , CD62L <sup>-</sup>

In order to determine ZEB2 expression across helper lineage populations from both Treg and Tconv, a combination of chemokine receptors was used following a modified gating strategy from Höllbacher et al. (2020) & Hope et al. (2019). For panel 3, four different chemokine receptors (CXCR3, CCR6, CCR4, CCR10) were used on CD45RA<sup>-</sup> memory Treg and memory Tconv to classify different helper lineage populations from both Treg and Tconv: Th1, Th2,

Th1/17, Th17, Th9 and Th22, Treg1, Treg2, Treg1/17, Treg17, Treg9 and Treg22 (Table 3.3). Enriched CD4<sup>+</sup> T cells were further pre-enriched with CD25 or CD45RA microbeads (Section 2.1.3), rested overnight in CX-VIVO medium. The next day, cells were then labelled with the following antibodies: BUV395 anti-human CD4, BV-421 anti-human CD25, PE-CF594 anti-human CD127, APC-H7 anti-human CD45RA, BV650 anti-human CXCR3, BV786 anti-human CCR6, Alexa Fluor® 647 anti-human CCR4, and BB515 anti-human CCR10 (Table 2.1).

**Table 3.3: Panel 3 – Chemokine receptors panel**

Cell type	Surface markers / Gating strategy
Th1	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>+</sup> , CCR6 <sup>-</sup>
Th2	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>-</sup> , CCR4 <sup>+</sup> , CCR10 <sup>-</sup>
Th1/17	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>+</sup> , CCR6 <sup>+</sup>
Th17	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>+</sup> , CCR10 <sup>-</sup>
Th22	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>+</sup> , CCR10 <sup>+</sup>
Th9	CD4 <sup>+</sup> , CD25 <sup>lo</sup> , CD127 <sup>hi</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>-</sup> , CCR10 <sup>-</sup>
Treg1	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>+</sup> , CCR6 <sup>-</sup>
Treg2	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>-</sup> , CCR4 <sup>+</sup> , CCR10 <sup>-</sup>
Treg1/17	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>+</sup> , CCR6 <sup>+</sup>
Treg17	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>+</sup> , CCR10 <sup>-</sup>
Treg22	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>+</sup> , CCR10 <sup>+</sup>
Treg9	CD4 <sup>+</sup> , CD25 <sup>hi</sup> , CD127 <sup>lo</sup> , CD45RA <sup>-</sup> , CXCR3 <sup>-</sup> , CCR6 <sup>+</sup> , CCR4 <sup>-</sup> , CCR10 <sup>-</sup>

Antigen-experienced T cells can be divided into CM and EM. In panel 4, we used CD62L to segregate CM and EM of Treg and Tconv helper lineage populations: Th1 CM, Th1 EM, Th2 CM, Th2 EM, Th1/17 CM, Th1/17 EM, Th17 CM, Th17 EM, Th22 CM, Th22 EM, Th9 CM, Th9 EM, Treg1 CM, Treg1 EM, Treg2 CM, Treg2 EM, Treg1/17 CM, Treg1/17 EM, Treg17 CM, Treg17 EM, Treg22 CM, Treg22 EM, Treg9 CM, Treg9 EM (Table 3.4). Enriched CD4<sup>+</sup> T cells were further pre-enriched with CD45RA microbeads (Section 2.1.3), rested overnight in CX-VIVO medium. The next day, cells were then labelled with following antibodies BUV395 anti-human CD4, BV-421 anti-human CD25, PE-CF594 anti-human CD127, APC-H7 anti-human CD45RA, BV650 anti-human CXCR3, BV786 anti-human CCR6, Alexa

Fluor® 647 anti-human CCR4, BB515 anti-human CCR10 and PE anti-human CD62L (Table 2.1).

**Table 3.4: Panel 4 – Memory chemokine receptor panel**

Cell type	Surface markers / Gating strategy
Th1 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3+, CCR6-, CD62L+
Th1 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3+, CCR6-, CD62L-
Th2 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6-, CCR4+, CCR10-, CD62L+
Th2 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6-, CCR4+, CCR10-, CD62L-
Th1/17 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3+, CCR6+, CD62L+
Th1/17 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3+, CCR6+, CD62L-
Th17 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10-, CD62L+
Th17 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10-, CD62L-
Th22 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10+, CD62L+
Th22 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10+, CD62L-
Th9 CM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4-, CCR10-, CD62L+
Th9 EM	CD4+, CD25lo, CD127hi, CD45RA-, CXCR3-, CCR6+, CCR4-, CCR10-, CD62L-
Treg1 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3+, CCR6-, CD62L+
Treg1 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3+, CCR6-, CD62L-
Treg2 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6-, CCR4+, CCR10-, CD62L+
Treg2 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6-, CCR4+, CCR10-, CD62L-
Treg1/17 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3+, CCR6+, CD62L+
Treg1/17 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3+, CCR6+, CD62L-
Treg17 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10-, CD62L+
Treg17 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10-, CD62L-
Treg22 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10+, CD62L+
Treg22 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4+, CCR10+, CD62L-
Treg9 CM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4-, CCR10-, CD62L+
Treg9 EM	CD4+, CD25hi, CD127lo, CD45RA-, CXCR3-, CCR6+, CCR4-, CCR10-, CD62L-

Cells labelled with antibodies (panel above) were sort purified with a FACS Fusion (BD Biosciences, San Jose, USA) flow cytometer (Section 2.1.6). CM and EM of Treg helper-like lineage populations were not sorted for panel 4, as EM Tregs were generally present in very low percentages in the periphery, therefore it was impractical to include these subsets for sort purification.

Sort purified cells were washed, pelleted and resuspended in appropriate medium for further analysis. Cells were checked for viability and counted as described above (Section 2.1.1).

Viable cells were then cultured or treated appropriately for further use. Cells sorted for RNA-

based gene expression analysis, were resuspended in 700 $\mu$ L of QIAzol<sup>®</sup> Lysis Reagent (QIAGEN, Germany) and mixed by vortexing for 15 seconds to lyse the cells, and then stored at -20°C for RNA extraction.

### **3.3.4 Detection of ZEB2 by Flow Cytometry in Treg and Tconv**

Detection of ZEB2 protein using flow cytometry allows quantitative analysis of its expression at the single-cell level. Therefore, when coupled with phenotypic surface markers, ZEB2 protein levels can be measured across different CD4<sup>+</sup> T cell subsets in a comprehensive way. ZEB2 RNA expression differences between isolated Treg and Tconv from panel 1 in Section 3.3.3 were first determined using qRT-PCR with a ZEB2 primer pair (

Table 2.3) following the protocol described in Section 2.2.4. To detect ZEB2 protein in Treg and Tconv, cells were first labelled with cell surface antibodies from Treg and Tconv panel 1 (Table 3.1) followed by nuclear protein labelling with either Alexa Fluor® 647 anti-human ZEB2, Alexa Fluor® 647 anti-human FOXP3 (Treg positive control) or Alexa Fluor® 647 anti-human IgG2B (negative control) (Table 3.5). Several controls were included in this experiment, cells labelled with FOXP3 antibody were used as a positive control for nuclear protein labelling, as FOXP3 is expressed more highly in Treg than in Tconv, IgG2B antibody labelling of cells was used as an isotype negative control to account for any non-specific background signals, surface stained alone and unstained controls were also included as controls for setting up physical gating (Figure 3.5).

Briefly,  $1 \times 10^6$  surface labelled cells were dispensed into each Falcon® FACS tube (Cat# 352008) according to Table 3.5 and including one tube for  $1 \times 10^6$  of unlabelled cells. Freshly prepared eBioscience Foxp3 Fixation/Permeabilisation buffer (1mL) was then added to each sample tube and mixed by vortexing. Sample tubes were incubated at 4°C for 30 minutes in the dark, and then washed twice with 2mL of eBioscience 1x Permeabilisation buffer. Cells were then incubated with the addition of either 5µL of anti-human ZEB2 conjugated with Alexa Fluor® 647 antibody (Cat#IC73782R, clone: 923328, R&D Systems, USA), 5µL of anti-human IgG2B conjugated with Alexa Fluor® 647 antibody (Cat# IC0041R, clone: 133303, R&D Systems, USA) or 20µL of anti-human FOXP3 conjugated with Alexa Fluor® 647 antibody (Cat# 560045, clone: 259D/C7, BD Biosciences, San Jose, USA) for 30 minutes at 4 °C in the dark. Samples were washed twice with 2mL of eBioscience 1x Permeabilisation buffer and once with 2mL PBS. Cells were then pelleted and resuspended in 200µL PBS and stored at 4 °C in the dark or analysed by flow cytometry (BD FACS Canto, BD Biosciences, San Jose, USA), and raw FCS data files were then analysed using FlowJo™ 10 software (FlowJo, LLC, USA).

**Table 3.5: Experimental FACS tubes for ZEB2 antibody labelling in Treg and Tconv**

Tubes	Antibodies
Unstained	None
Surface stained only	CD4, CD25, CD127
Surface stained + IgG2B	CD4, CD25, CD127, IgG2B
Surface stained + FOXP3	CD4, CD25, CD127, FOXP3
Surface stained + ZEB2	CD4, CD25, CD127, ZEB2

### 3.3.5 UMAP Dimensionality Reduction of T Helper Lineage Populations

Uniform Manifold Approximation and Projection (UMAP) is a machine learning algorithm used for dimensionality reduction to visualize high parameter datasets in a two-dimensional space. The UMAP plugin (Becht et al., 2018) was run on FlowJo™ 10 software (FlowJo, LLC, USA). First, raw FCS data from enriched CD4<sup>+</sup> T cells labelled with surface antibodies from panel 4 were imported into FlowJo™ 10 software (FlowJo, LLC, USA). Any anomalies in the raw data were removed with FlowAI plugin (Monaco et al., 2016) and then memory Tconv were then physically gated. Event numbers of memory Tconv were downsampled to the same number across donors followed by concatenation. UMAP dimension reduction was then performed with default settings on concatenated data. The number of clusters from the UMAP were determined by XShift plugin (Samusik et al., 2016) using default settings and the predicted number of clusters was then added as a parameter into FlowSom plugin (Van Gassen et al., 2015) to obtain a clear and unbiased overview of subsets from memory Tconv. These detected clusters were applied on UMAP graph layout and illustrated using the ClusterExplorer plugin (Trotter, 2019).

### 3.3.6 Central and Effector Memory Helper Lineage Activation

To understand the kinetics of ZEB2 expression following activation in CM and EM helper lineage populations, RNA was collected from  $5 \times 10^4$  sort purified Th1 CM, Th1 EM, Th2 CM,

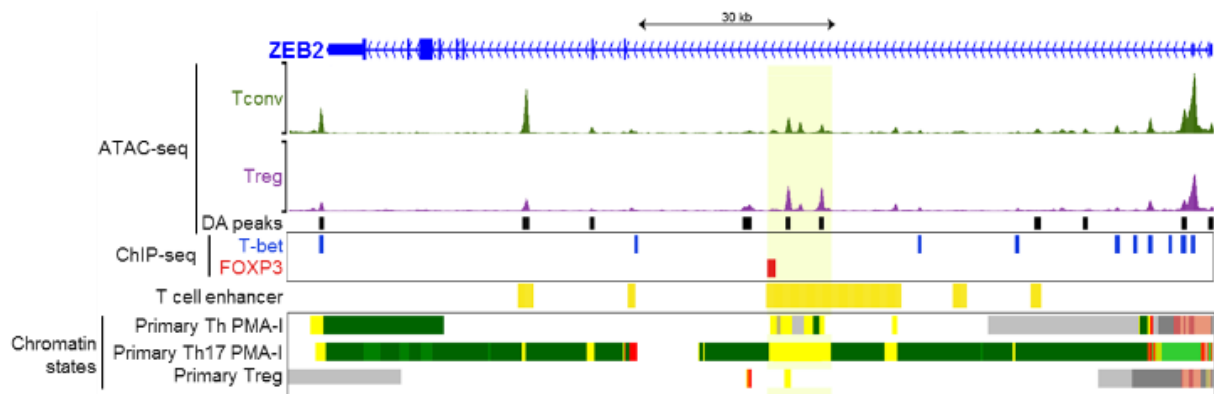


Th2 EM, Th1/17 CM and Th1/17 EM cells (Table 3.4) on the day of isolation and 5 days post stimulation. Cells were activated by stimulation with anti-CD3/CD28 coated magnetic beads (1:1 anti-CD3/CD28 magnetic bead, Dynabeads® Human T-Activator CD3/CD28, Life Technologies) in the presence of IL-2 (100 Units/mL) following methods in Section 2.1.7. ZEB2, T-bet, GATA3 and ROR $\gamma$ t mRNA expression kinetics were measured using qRT-PCR (Section 2.2.3 & 2.2.4) on samples collected from day 0 and day 5.

## 3.4 Results

### 3.4.1 ZEB2 Gene Locus Accessibility

Chromatin accessibility profiling of the ZEB2 locus (Figure 3.1) showed increased chromatin accessibility in Tconv compared with Tregs cells (Wong et al., unpublished observations). This is unsurprising and further supports the expression differences of ZEB2 mRNA in Treg and Tconv, when our lab first identified this gene as differentially expressed in Treg and Tconv. Interestingly, broadcast of T-bet ChIP-seq data (Hertweck et al., 2016) on the ZEB2 locus shows T-bet binding sites in the ZEB2 locus intersecting with regions of increased chromatin accessibility in Tconv (Figure 3.1). I also noted that there is a strong FOXP3 binding site in the ZEB2 locus in intron 2, which was previously validated by ChIP-PCR and reporter assay (Brown et al., 2018). Interestingly, the region containing the FOXP3 binding site is also in proximity with an enhancer region that showed increased chromatin accessibility in Treg compared with Tconv. Together, these tracks suggest that ZEB2 transcription is repressed by FOXP3 in Treg, but potentially induced in CD4 cells by T-bet.



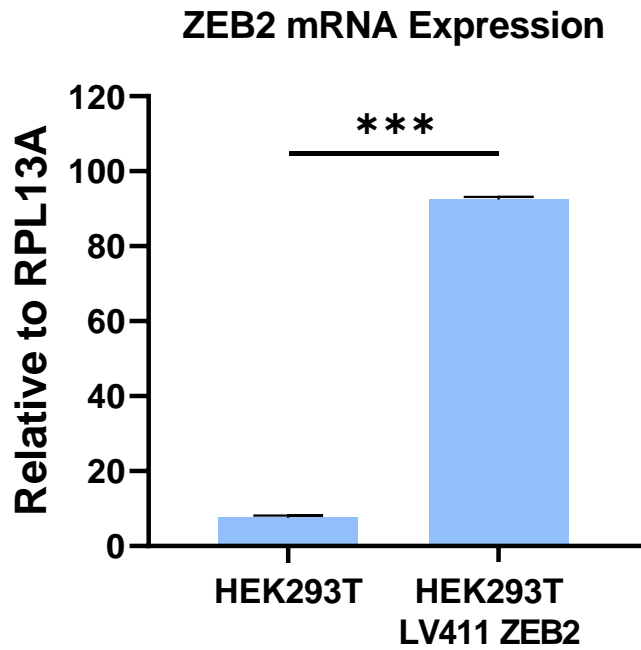
**Figure 3.1: Visualizing of ZEB2 gene locus of Treg and Tconv**

ATAC-seq tracks show chromatin accessibility at the ZEB2 gene locus of Treg (purple) and Tconv (green) and differentially accessible peaks (black) (Wong et al., unpublished) with FOXP3 ChIP (red) (Sadlon et al., 2010), T-bet ChIP (blue) (Hertweck et al., 2016), enhancer annotation (yellow) and chromatin states from NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al., 2010).

### 3.4.2 Detection of ZEB2 Protein Expression by Flow Cytometry

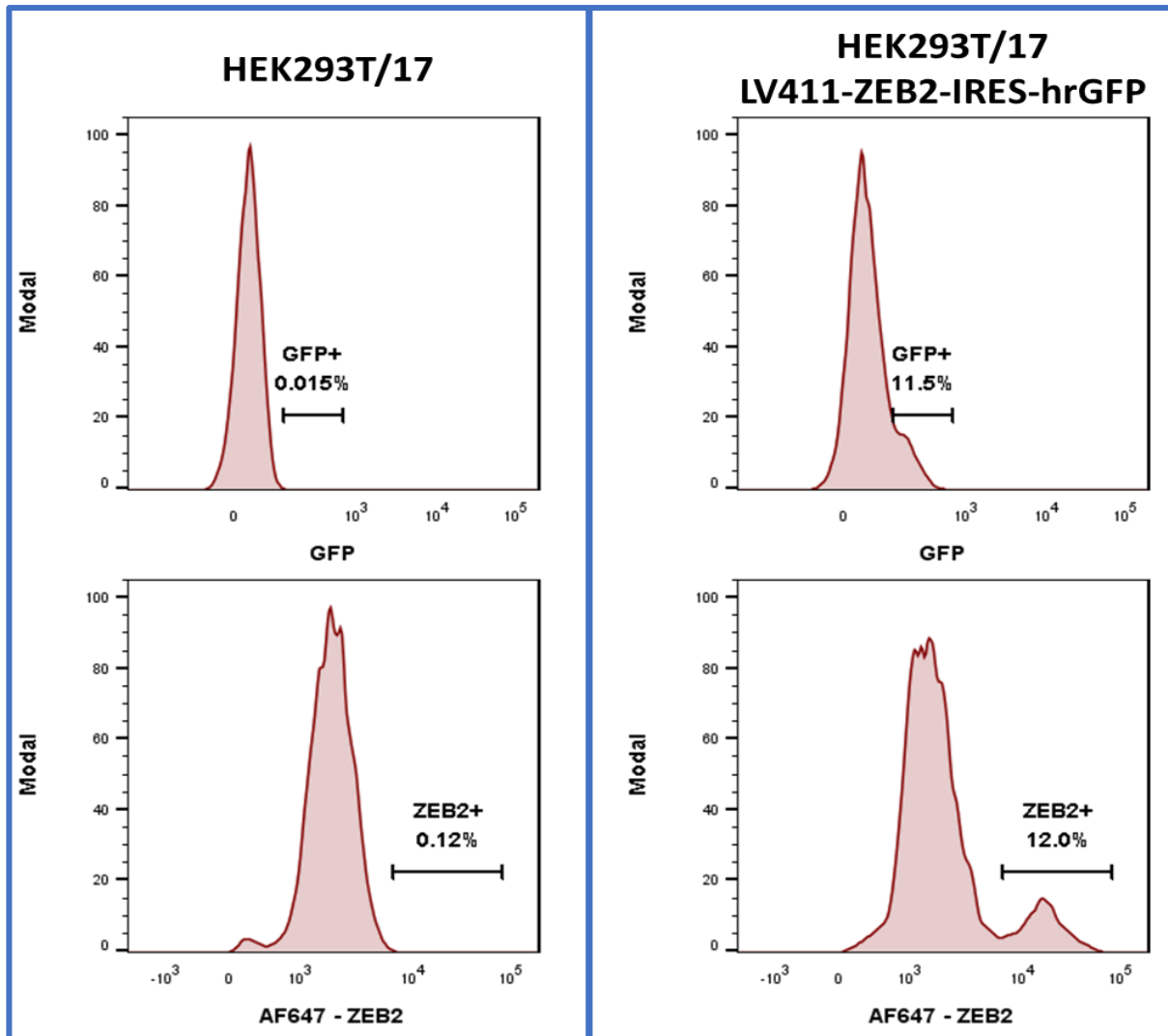
#### ZEB2 antibody labelling in HEK293T/17 cell lines overexpressing ZEB2

The expression of ZEB2 protein can be identified by flow cytometry in the populations of CD4<sup>+</sup> T cells. Currently, there is no working antibody that specifically detect human ZEB2 protein by flow cytometry in CD4<sup>+</sup> T cell, therefore I first tested an anti-ZEB2 antibody (Alexa Fluor® 647 conjugated anti-human ZEB2/SIP1 antibody, clone 923328, R&D Systems, USA) using the HEK293T/17 cell line transduced with LV411-ZEB2-IRES-hrGFP, to overexpress ZEB2 (vector maps in Figure 2.8). Prior to antibody staining, I first quantitated expression of ZEB2 mRNA in both un-transduced HEK392T/17 cells and in cells transduced with LV-411-ZEB2-IRES-GFP using qRT-PCR (Figure 3.2). Expression of ZEB2 mRNA was >80x greater in ZEB2 overexpressing HEK293T/17 in comparison with the untransduced HEK293T/17 cells, confirming that ZEB2 was indeed overexpressed. Cells were then labelled with the anti-ZEB2 antibody (anti-human ZEB2/SIP1 antibody, clone 923328, R&D Systems, USA, 5µL per 1x10<sup>6</sup> cells) and analysed by flow cytometry (FACS Canto, BD Biosciences, San Jose, USA). Approximately 11-12% of the LV-411-ZEB2-IRES-GFP transduced HEK293T/17 cells were GFP<sup>+</sup>, and likewise, 11-12% of LV-411-ZEB2-IRES-GFP transduced HEK293T/17 cells were ZEB2<sup>+</sup>, it is highly likely that these are the same cell populations (Figure 3.3). As it is widely known that permeabilising reagents can quench GFP signals in a cell, it was therefore not feasible to accurately validate the co-expression of GFP and AF-647 signal. ZEB2 protein detection by flow cytometry was observed in 3 independent experiments. Overall, significantly increased of ZEB2 mRNA and protein were measured in ZEB2-overexpressing HEK293T/17 cells compared with the control HEK293T/17 cells. In addition, antibody labelling of ZEB2 protein for flow cytometry was also validated so that it could be used for phenotyping ZEB2 expression across various CD4<sup>+</sup> T cell subsets.



**Figure 3.2: ZEB2 mRNA expression in HEK293T & HEK293T overexpressing ZEB2**

ZEB2 expression in WT or ZEB2 overexpressing HEK293T cells. Relative abundance of ZEB2 mRNA normalised to reference gene RPL13A and plotted with mean + SEM, Paired t test, \*\*\* $p < 0.001$  n = 3 independent experiments.

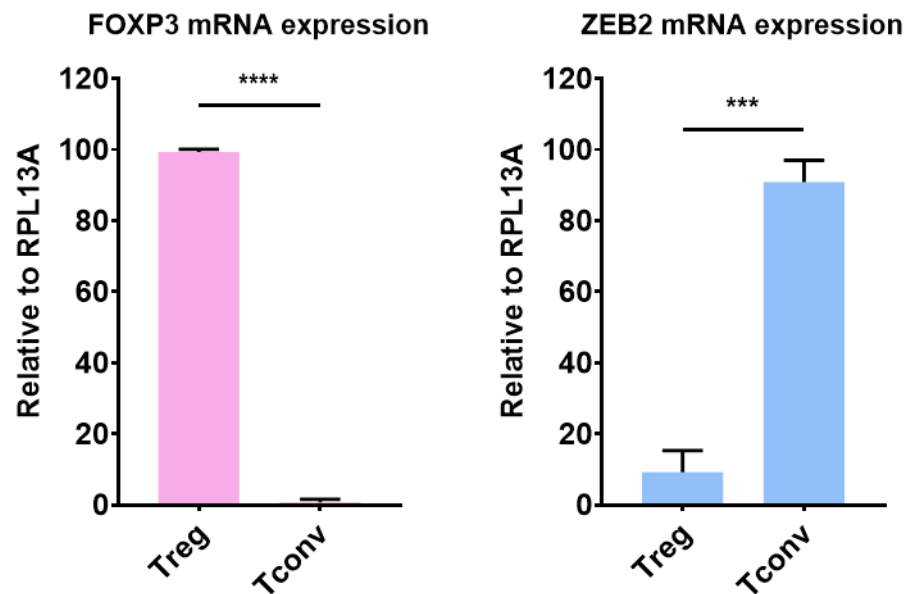


**Figure 3.3: Analysis of ZEB2 protein level in HEK293T & HEK293T overexpressing ZEB2**

Histogram showing ZEB2 protein level in ZEB2 overexpressing (LV411-ZEB2-IRES-hrGFP) HEK293T/17 cells (right) compared with control cells (left), both cell pools were labelled with anti-ZEB2 antibody. Representative data from 1 experiment, n=3 independent experiments.

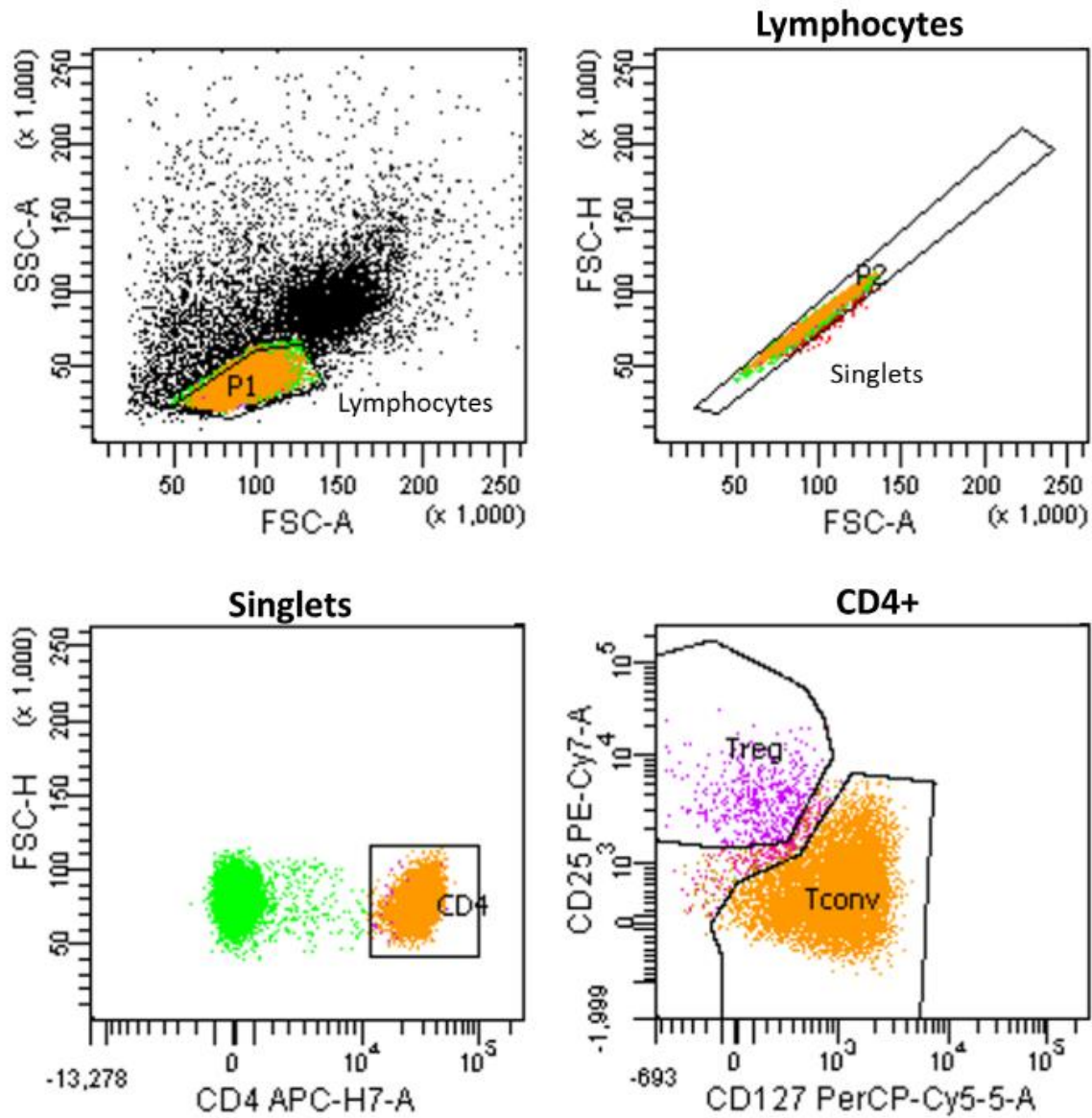
### ZEB2 antibody labelling in Treg and Tconv

Next, I wanted to confirm that ZEB2 protein could also be detected using the same ZEB2 antibody in the CD4<sup>+</sup> T cell compartment. FOXP3 and ZEB2 mRNA expression were first measured and these results confirmed our previous findings, that FOXP3 mRNA expression in Treg is higher than in Tconv and, conversely, that there is significantly higher ZEB2 mRNA expression in Tconv compared with Treg (Figure 3.4). To determine ZEB2 and FOXP3 protein expression in Tconv compared with Treg (Figure 3.4). To determine ZEB2 and FOXP3 protein expression, cells were first segregated into Treg (CD25<sup>hi</sup>, CD127<sup>lo</sup>) and Tconv CD25<sup>lo</sup>, CD127<sup>hi</sup>) from bulk CD4<sup>+</sup> T cell (Figure 3.5) using surface labelling antibodies in panel 1 as indicated in Section 3.3.3, followed by intracellular staining (see Methods in Section 3.3.4) using either the anti-ZEB2 antibody, or the anti-FOXP3 antibody as a positive control in Treg, or anti-IgG2B Ab as a negative isotype control, or no intracellular antibody as background control (Figure 3.6). Similar to FOXP3 mRNA found in Treg, its protein level can also be detected in Treg using an anti-FOXP3 antibody. Unfortunately, although FOXP3 was clearly detected, I could not detect ZEB2 protein in either Treg or Tconv even though I could clearly measure ZEB2 mRNA (Figure 3.4). This may be owing to a relatively low abundance of endogenous ZEB2 protein in CD4<sup>+</sup> T cell subsets such that the anti-ZEB2 antibody was not sensitive enough to detect. This inability to detect endogenous ZEB2 protein contrasts with the easy detection of ZEB2 protein in HEK293T/17 cells overexpressing ZEB2, but it must be remembered that in this scenario ZEB2 is artificially high and therefore it is easily detectable using the anti-ZEB2 antibody. These data suggested that ZEB2 protein could not be detected with antibody in CD4<sup>+</sup> T cells.



**Figure 3.4: FOXP3 and ZEB2 mRNA expression in Treg and Tconv**

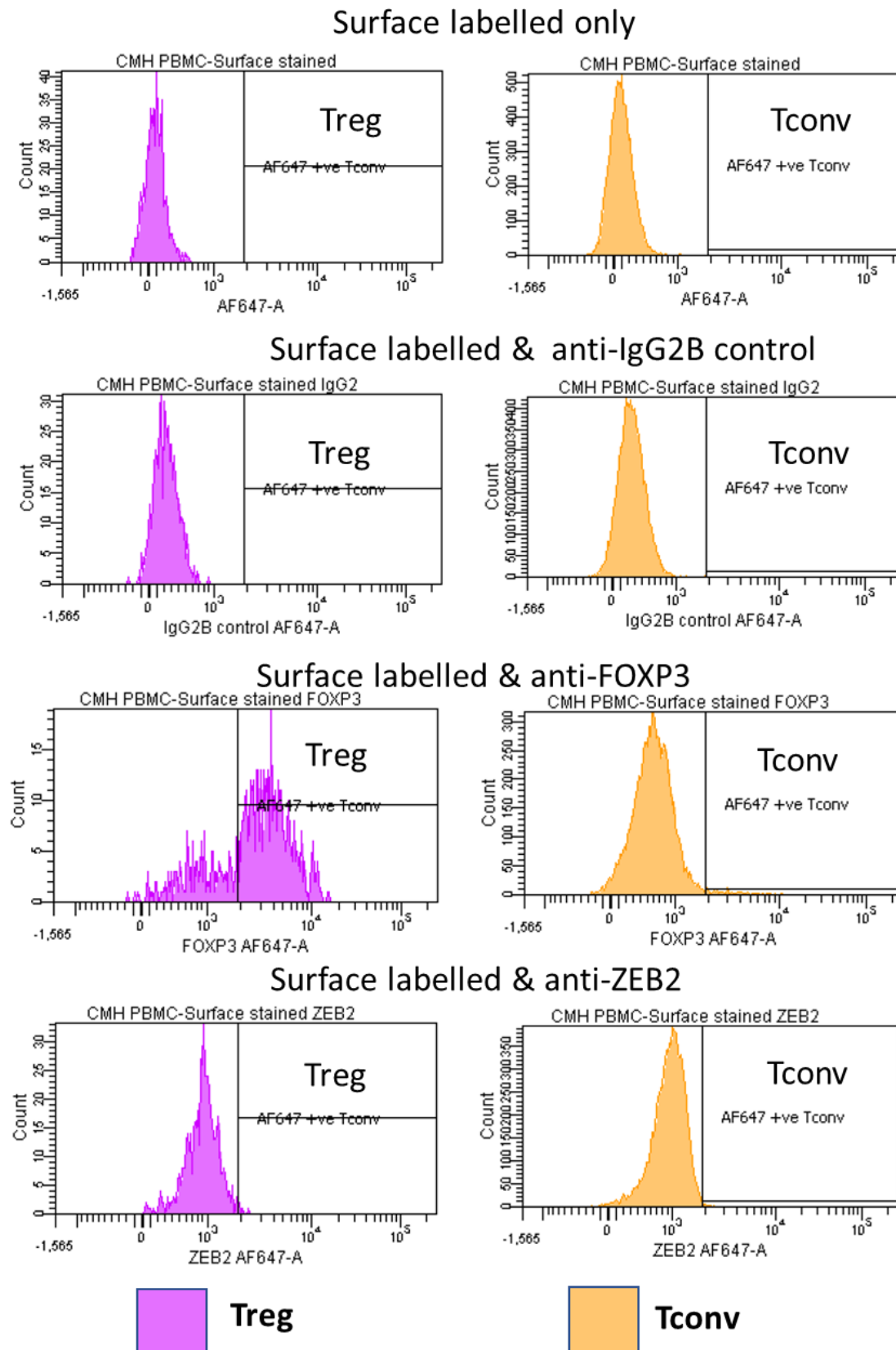
FOXP3 and ZEB2 expression Treg and Tconv. Relative abundance of (A) FOXP3 and (B) ZEB2 mRNA normalised to reference gene RPL13A and plotted with mean + SEM, Paired t test, \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ,  $n = 3$  independent donors.



**Figure 3.5: Physical gating strategy for Treg and Tconv**

Flow cytometry plot showing PBMC surface labelled with panel 1: CD4, CD25, CD127. Cells were first gated on lymphocytes, singlets, CD4+ before separating into Treg (CD25<sup>hi</sup>, CD127<sup>lo</sup>) and Tconv (CD25<sup>lo</sup>, CD127<sup>hi</sup>). Representative data from 1 donor, n= 3 independent donors.





**Figure 3.6: ZEB2 protein level in Treg and Tconv**

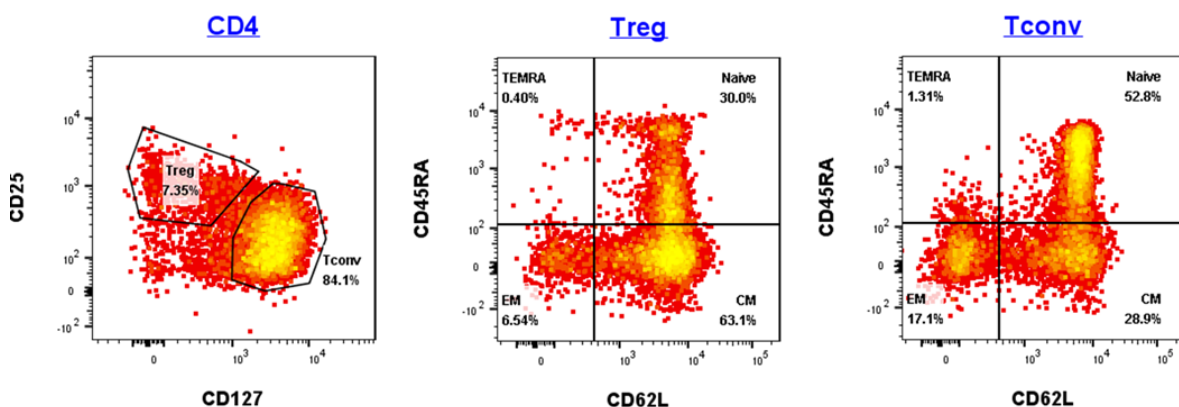
Histogram showing gated Treg (pink) and Tconv (orange) from Figure 3.5 intracellularly labelled with either no antibody (background), anti-IgG2B antibody (isotype negative control), anti-FOXP3 antibody (positive control for Treg) or anti-ZEB2 antibody. Representative data from 1 donor, n=3 independent donors.

A considerable amount of time was spent testing reagents and protocols for reliable and specific detection and measurement of ZEB2 protein in CD4<sup>+</sup> T cells by flow cytometry. I tried to optimise antibody labelling of both conjugated anti-ZEB2 antibody (Alexa Fluor® 647 conjugated anti-human ZEB2/SIP1 antibody, clone 923328, R&D Systems, USA) and a different manufacturer's non-conjugated anti-ZEB2 antibody (anti-SIP1/ZEB2 antibody, clone 6E5, Cat#61095, Active Motif) which I conjugated myself with a fluorophore kit (Alexa Fluor™ 647 Antibody Labelling Kit, Cat#A20186, Thermo Fisher Scientific Inc.). The reason for using this anti-ZEB2 antibody from Active motif was that I have previously successfully detected and measured endogenous ZEB2 protein expression in breast cancer cell lines and primary breast epithelia by western blot, and the Active Motif anti-ZEB2 antibody was the only antibody that clearly and specifically detected ZEB2 in these cells (Brown et al., 2018). Additionally, I also tried different techniques for fixing and permeabilising CD4<sup>+</sup> T cells to potentially allow for better antibody access to the cell. However, none of these protocols showed an improvement compared with the result in Figure 3.6, above, using the conjugated anti-ZEB2 antibody (Alexa Fluor® 647 conjugated anti-human ZEB2/SIP1 antibody, clone 923328, R&D Systems, USA) in the CD4<sup>+</sup> T cells and therefore the optimisation data and methods involved are not shown.

### **3.4.3 ZEB2 is Highly Expressed in Effector Memory CD4<sup>+</sup> T Cells**

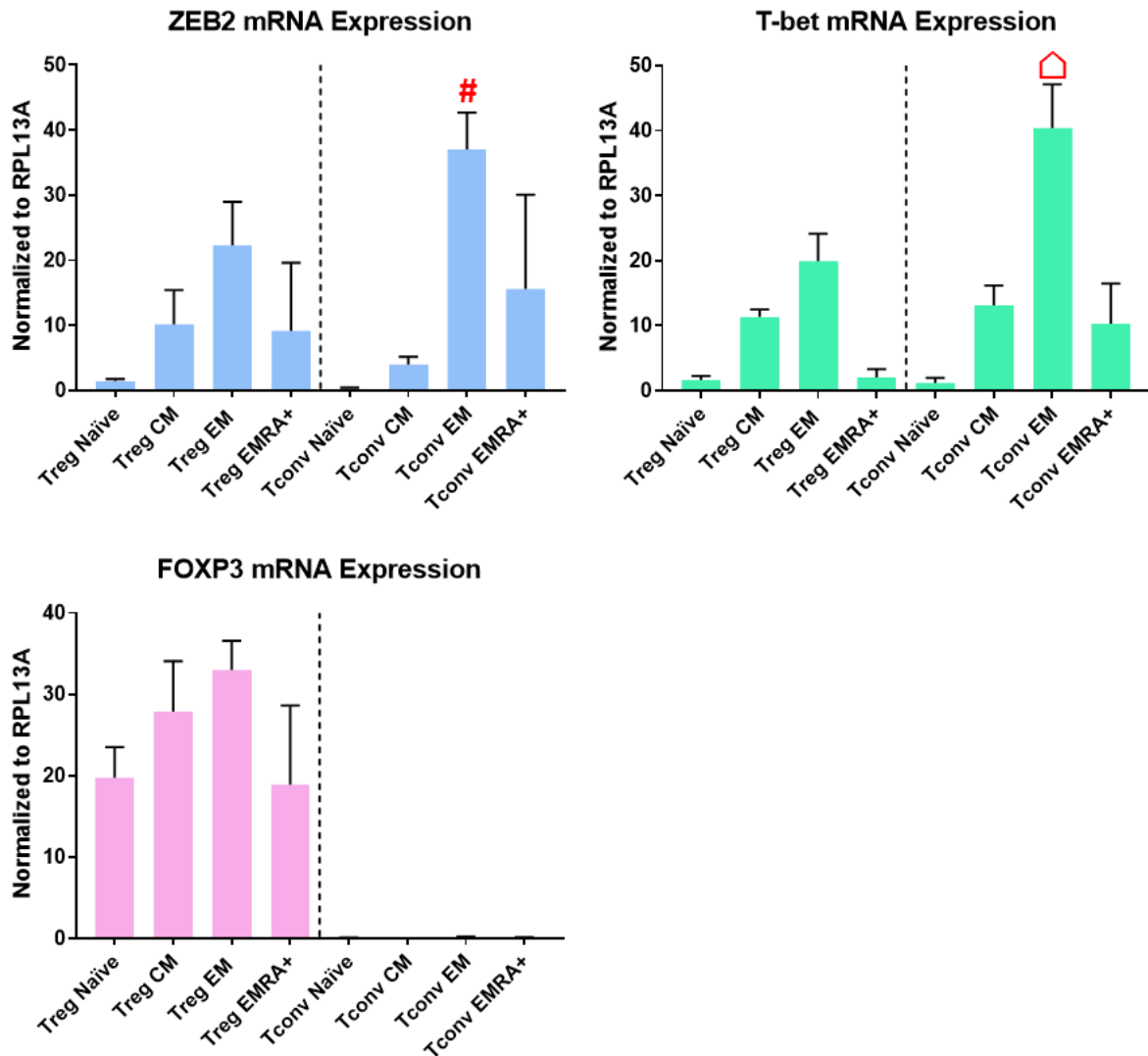
Having shown that ZEB2 protein detection by flow cytometry was not feasible with reagents currently available, from this point onwards all ZEB2 expression analysis was carried out using mRNA isolated from the CD4<sup>+</sup> subsets. In order to identify the CD4<sup>+</sup> T cell populations that express ZEB2, Treg and Tconv populations were gated and further subdivided into respective naïve, central memory (CM), effector memory (EM) and effector memory CD45RA<sup>+</sup> (emRA) with two surface markers CD45RA and CD62L respectively as indicated in panel 2 of Section 3.3.3 and Table 3.2 (Figure 3.7) and RNA was isolated from each of the populations (see

methods in Section 2.2.1). Expression of FOXP3, ZEB2 and T-bet mRNAs was then examined in each of the memory populations of Treg and Tconv. As we would predict, FOXP3 mRNA was expressed highly across all of the Treg memory populations but was very low in all of the Tconv populations (Figure 3.8). The expression profile of ZEB2 was quite interesting and very specific. ZEB2 expression was very low in naïve cells, slightly higher in the CM cell population and highest in the EM population. Although overall ZEB2 is more highly expressed in Tconv compared with Treg (see Figure 3.4 above), this pattern of expression across the maturation populations was observed in both Tconv and Treg. Expression of ZEB2 mRNA in the very rare TEMRA population was low but more variable (Figure 3.8), suggestive of heterogeneity of the TEMRA population between donors. Notably, the T-bet expression pattern was quite similar to that of ZEB2 expression in bulk CD4<sup>+</sup> T cell across memory subsets. Overall, these data suggest that ZEB2 may be important either in the effector function or in the differentiation of CD4<sup>+</sup> T Effector Memory cells.



**Figure 3.7: Memory subset profiling of Treg and Tconv**

Flow cytometry plot showing RosetteSep™ enriched CD4<sup>+</sup> surface stained with CD4, CD25, CD127, CD45RA, CD62L. Cells were first gated on lymphocytes, singlets, CD4<sup>+</sup> before separating into Treg (CD25<sup>hi</sup>, CD127<sup>lo</sup>) and Tconv (CD25<sup>lo</sup>, CD127<sup>hi</sup>). Naïve (CD45RA<sup>+</sup>, CD62L<sup>+</sup>), central memory (CM; CD45RA<sup>-</sup>, CD62L<sup>+</sup>), effector memory (EM; CD45RA<sup>-</sup>, CD462L<sup>-</sup>), effector memory RA<sup>+</sup> (emRA; CD45RA<sup>+</sup>, CD62L<sup>-</sup>) population were separated from gated Treg and Tconv. Representative data from 1 donor, n=3 independent donors.



**Figure 3.8: ZEB2, T-bet and FOXP3 mRNA expression in memory subset profile of Treg and Tconv**

ZEB2 (blue), T-bet (green) and FOXP3 (pink) expression in memory subset profile of Treg and Tconv. Relative abundance of ZEB2, T-bet and FOXP3 were normalised to reference gene RPL13A and plotted with mean + SEM. As there are multiple comparisons, the statistical significance is presented in a separate table, to keep the figures clear. The annotation # indicates the level of ZEB2 in Tconv EM is significantly higher than other subsets whereas the annotation  $\triangle$  indicates the level of T-bet in Tconv EM is significantly higher than other subsets. Statistics were carried out with ordinary one-way Anova multiple comparison Fisher's Least Significant Difference (LSD) test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$  as shown in Table 3.6, Table 3.7 and Table 3.8,  $n = 3$  independent donors.

**Table 3.6: Anova Fisher's LSD test on ZEB2 expression in memory subset profile of Treg and Tconv**

<b>Fisher's LSD test</b>	<b>Mean Diff.</b>	<b>95.00% CI of diff.</b>	<b>Summary</b>	<b>Individual P Value</b>
<i>Treg Naïve vs. Treg CM</i>	-8.701	-24.14 to 6.740	ns	0.2497
<i>Treg Naïve vs. Treg EM</i>	-20.82	-36.26 to -5.375	*	0.0114
<i>Treg Naïve vs. Treg EMRA+</i>	-7.707	-23.15 to 7.734	ns	0.3057
<i>Treg Naïve vs. Tconv Naïve</i>	1.189	-14.25 to 16.63	ns	0.8724
<i>Treg Naïve vs. Tconv CM</i>	-2.523	-17.96 to 12.92	ns	0.7335
<i>Treg Naïve vs. Tconv EM</i>	-35.57	-51.01 to -20.13	***	0.0002
<i>Treg Naïve vs. Tconv EMRA+</i>	-14.15	-29.59 to 1.290	ns	0.0698
<i>Treg CM vs. Treg EM</i>	-12.12	-27.56 to 3.326	ns	0.1157
<i>Treg CM vs. Treg EMRA+</i>	0.9939	-14.45 to 16.43	ns	0.8932
<i>Treg CM vs. Tconv Naïve</i>	9.889	-5.551 to 25.33	ns	0.1934
<i>Treg CM vs. Tconv CM</i>	6.177	-9.263 to 21.62	ns	0.4089
<i>Treg CM vs. Tconv EM</i>	-26.87	-42.31 to -11.43	**	0.0020
<i>Treg CM vs. Tconv EMRA+</i>	-5.450	-20.89 to 9.991	ns	0.4651
<i>Treg EM vs. Treg EMRA+</i>	13.11	-2.332 to 28.55	ns	0.0908
<i>Treg EM vs. Tconv Naïve</i>	22.00	6.564 to 37.45	**	0.0081
<i>Treg EM vs. Tconv CM</i>	18.29	2.852 to 33.73	*	0.0231
<i>Treg EM vs. Tconv EM</i>	-14.75	-30.20 to 0.6862	ns	0.0598
<i>Treg EM vs. Tconv EMRA+</i>	6.665	-8.776 to 22.11	ns	0.3738
<i>Treg EMRA+ vs. Tconv Naïve</i>	8.896	-6.545 to 24.34	ns	0.2397
<i>Treg EMRA+ vs. Tconv CM</i>	5.184	-10.26 to 20.62	ns	0.4869
<i>Treg EMRA+ vs. Tconv EM</i>	-27.86	-43.30 to -12.42	**	0.0015
<i>Treg EMRA+ vs. Tconv EMRA+</i>	-6.444	-21.89 to 8.997	ns	0.3894
<i>Tconv Naïve vs. Tconv CM</i>	-3.712	-19.15 to 11.73	ns	0.6173
<i>Tconv Naïve vs. Tconv EM</i>	-36.76	-52.20 to -21.32	***	0.0001
<i>Tconv Naïve vs. Tconv EMRA+</i>	-15.34	-30.78 to 0.1011	ns	0.0513
<i>Tconv CM vs. Tconv EM</i>	-33.05	-48.49 to -17.61	***	0.0003
<i>Tconv CM vs. Tconv EMRA+</i>	-11.63	-27.07 to 3.813	ns	0.1300
<i>Tconv EM vs. Tconv EMRA+</i>	21.42	5.979 to 36.86	**	0.0096

**Table 3.7: Anova Fisher's LSD test on T-bet expression in memory subset profile of Treg and Tconv**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg Naïve vs. Treg CM</i>	-9.751	-19.66 to 0.1564	ns	0.0533
<i>Treg Naïve vs. Treg EM</i>	-18.34	-28.25 to -8.432	**	0.0012
<i>Treg Naïve vs. Treg EMRA+</i>	-0.4091	-10.32 to 9.498	ns	0.9313
<i>Treg Naïve vs. Tconv Naïve</i>	0.3946	-9.513 to 10.30	ns	0.9338
<i>Treg Naïve vs. Tconv CM</i>	-11.50	-21.41 to -1.597	*	0.0256
<i>Treg Naïve vs. Tconv EM</i>	-38.78	-48.69 to -28.87	****	<0.0001
<i>Treg Naïve vs. Tconv EMRA+</i>	-8.700	-18.61 to 1.208	ns	0.0812
<i>Treg CM vs. Treg EM</i>	-8.588	-18.50 to 1.319	ns	0.0848
<i>Treg CM vs. Treg EMRA+</i>	9.342	-0.5655 to 19.25	ns	0.0629
<i>Treg CM vs. Tconv Naïve</i>	10.15	0.2382 to 20.05	*	0.0453
<i>Treg CM vs. Tconv CM</i>	-1.753	-11.66 to 8.154	ns	0.7125
<i>Treg CM vs. Tconv EM</i>	-29.03	-38.94 to -19.12	****	<0.0001
<i>Treg CM vs. Tconv EMRA+</i>	1.051	-8.856 to 10.96	ns	0.8249
<i>Treg EM vs. Treg EMRA+</i>	17.93	8.023 to 27.84	**	0.0015
<i>Treg EM vs. Tconv Naïve</i>	18.73	8.826 to 28.64	**	0.0010
<i>Treg EM vs. Tconv CM</i>	6.835	-3.072 to 16.74	ns	0.1630
<i>Treg EM vs. Tconv EM</i>	-20.44	-30.35 to -10.53	***	0.0005
<i>Treg EM vs. Tconv EMRA+</i>	9.640	-0.2679 to 19.55	ns	0.0558
<i>Treg EMRA+ vs. Tconv Naïve</i>	0.8037	-9.104 to 10.71	ns	0.8656
<i>Treg EMRA+ vs. Tconv CM</i>	-11.09	-21.00 to -1.187	*	0.0305
<i>Treg EMRA+ vs. Tconv EM</i>	-38.37	-48.28 to -28.46	****	<0.0001
<i>Treg EMRA+ vs. Tconv EMRA+</i>	-8.291	-18.20 to 1.617	ns	0.0951
<i>Tconv Naïve vs. Tconv CM</i>	-11.90	-21.81 to -1.991	*	0.0216
<i>Tconv Naïve vs. Tconv EM</i>	-39.17	-49.08 to -29.27	****	<0.0001
<i>Tconv Naïve vs. Tconv EMRA+</i>	-9.094	-19.00 to 0.8132	ns	0.0694
<i>Tconv CM vs. Tconv EM</i>	-27.27	-37.18 to -17.37	****	<0.0001
<i>Tconv CM vs. Tconv EMRA+</i>	2.804	-7.103 to 12.71	ns	0.5569
<i>Tconv EM vs. Tconv EMRA+</i>	30.08	20.17 to 39.99	****	<0.0001

**Table 3.8: Anova Fisher's LSD test on FOXP3 expression in memory subset profile of Treg and Tconv**

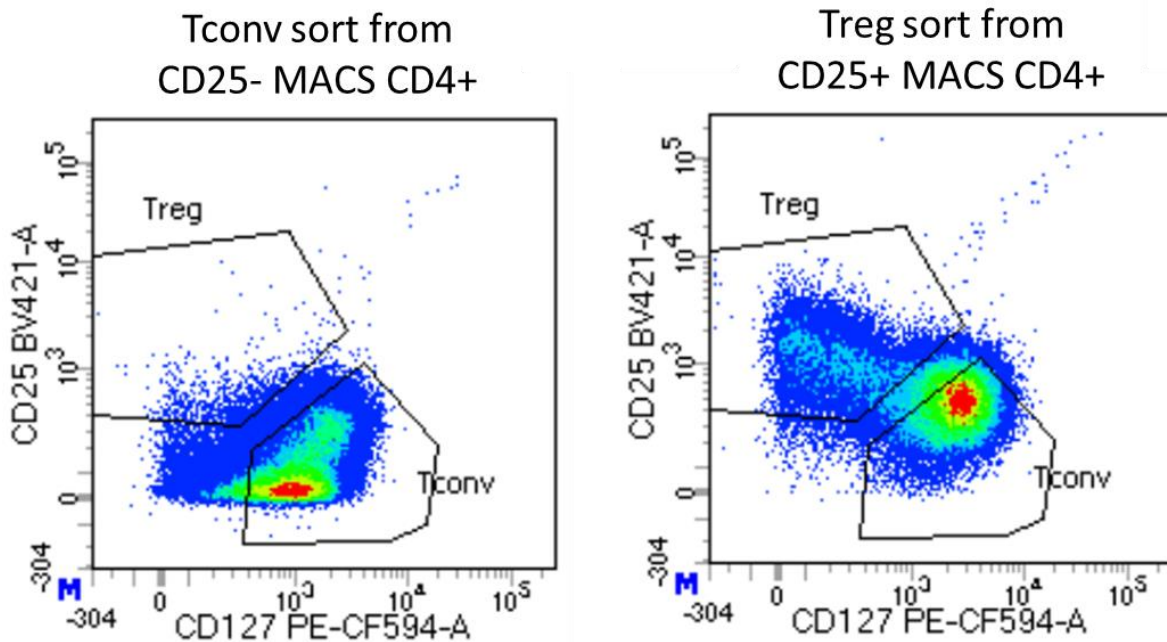
Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg Naïve vs. Treg CM</i>	-8.152	-18.56 to 2.260	ns	0.1164
<i>Treg Naïve vs. Treg EM</i>	-13.27	-23.68 to -2.853	*	0.0157
<i>Treg Naïve vs. Treg EMRA+</i>	0.8379	-9.574 to 11.25	ns	0.8667
<i>Treg Naïve vs. Tconv Naïve</i>	19.67	9.262 to 30.09	**	0.0010
<i>Treg Naïve vs. Tconv CM</i>	19.62	9.208 to 30.03	**	0.0010
<i>Treg Naïve vs. Tconv EM</i>	19.56	9.147 to 29.97	**	0.0011
<i>Treg Naïve vs. Tconv EMRA+</i>	19.63	9.219 to 30.04	**	0.0010
<i>Treg CM vs. Treg EM</i>	-5.113	-15.53 to 5.299	ns	0.3133
<i>Treg CM vs. Treg EMRA+</i>	8.990	-1.422 to 19.40	ns	0.0859
<i>Treg CM vs. Tconv Naïve</i>	27.83	17.41 to 38.24	****	<0.0001
<i>Treg CM vs. Tconv CM</i>	27.77	17.36 to 38.18	****	<0.0001
<i>Treg CM vs. Tconv EM</i>	27.71	17.30 to 38.12	****	<0.0001
<i>Treg CM vs. Tconv EMRA+</i>	27.78	17.37 to 38.20	****	<0.0001
<i>Treg EM vs. Treg EMRA+</i>	14.10	3.691 to 24.52	*	0.0111
<i>Treg EM vs. Tconv Naïve</i>	32.94	22.53 to 43.35	****	<0.0001
<i>Treg EM vs. Tconv CM</i>	32.89	22.47 to 43.30	****	<0.0001
<i>Treg EM vs. Tconv EM</i>	32.82	22.41 to 43.24	****	<0.0001
<i>Treg EM vs. Tconv EMRA+</i>	32.90	22.48 to 43.31	****	<0.0001
<i>Treg EMRA+ vs. Tconv Naïve</i>	18.84	8.424 to 29.25	**	0.0015
<i>Treg EMRA+ vs. Tconv CM</i>	18.78	8.370 to 29.19	**	0.0015
<i>Treg EMRA+ vs. Tconv EM</i>	18.72	8.309 to 29.13	**	0.0015
<i>Treg EMRA+ vs. Tconv EMRA+</i>	18.79	8.381 to 29.21	**	0.0015
<i>Tconv Naïve vs. Tconv CM</i>	-0.05364	-10.47 to 10.36	ns	0.9914
<i>Tconv Naïve vs. Tconv EM</i>	-0.1145	-10.53 to 10.30	ns	0.9817
<i>Tconv Naïve vs. Tconv EMRA+</i>	-0.04305	-10.46 to 10.37	ns	0.9931
<i>Tconv CM vs. Tconv EM</i>	-0.06083	-10.47 to 10.35	ns	0.9903
<i>Tconv CM vs. Tconv EMRA+</i>	0.01059	-10.40 to 10.42	ns	0.9983
<i>Tconv EM vs. Tconv EMRA+</i>	0.07143	-10.34 to 10.48	ns	0.9886

### 3.4.4 ZEB2 Expression in CD25 MACS Enriched Purified Treg and Tconv Helper Lineages

In order to respond to the broad range of environmental challenges encountered by CD4+ Effector T cells, these cells can differentiate into populations that express a range of effector molecules specific for particular types of antigenic challenge (see Section 1.4 for more detail). Since Treg specifically suppress an effector response, it is now considered highly likely that Treg likewise differentiate into T-effector-targeting subsets that specifically target their T effector counterpart (Shevyrev and Tereshchenko, 2020). These different memory helper lineage populations (or subsets) of cells can be isolated based on their expression of particular chemokine receptors (Duhon et al., 2012). To capture the extent of this heterogeneity within the memory compartment of Tconv and Treg, I modified the flow cytometry gating strategy of Hope et al. (2019) and Höllbacher et al. (2020) to isolate various helper subsets based on chemokine receptors expressed on the cells. The following chemokine receptors were used for these studies: CXCR3, CCR4, CCR6, CCR10. In order to reduce sorting time and increase sorting efficiency for resolving target cells from non-target cells, CD25 MACS was used to enrich for Treg in the CD25+ fraction and Tconv in the CD25- fraction (Figure 3.9). The CD25- fraction was depleted of CD25hi Tconv cells and enriched for CD25- Tconv and the CD25+ fraction enriched for Treg cells and contained some of the CD25hi Tconv cells. Gated memory Tconv were then separated into different helper lineages: Th1 (CXCR3+, CCR6-), Th2 (CXCR3-, CCR6-), Th1/17 (CXCR3+, CCR6+), Th9/17 (CXCR3-, CCR6+, CCR10-) and Th22 (CXCR3-, CCR6+, CCR10+, CCR4+) and the gated memory Treg were similarly separated into Treg1 (CXCR3+, CCR6-), Treg2 (CXCR3-, CCR6-), Treg1/17 (CXCR3+, CCR6+), Treg9/17 (CXCR3-, CCR6+, CCR10-) and Treg22 (CXCR3-, CCR6+, CCR10+, CCR4+) (Figure 3.10). In this experiment, Treg9 and Treg17 were not separated as some donors have very few Treg9. Therefore, it was not feasible to isolate these two populations separately

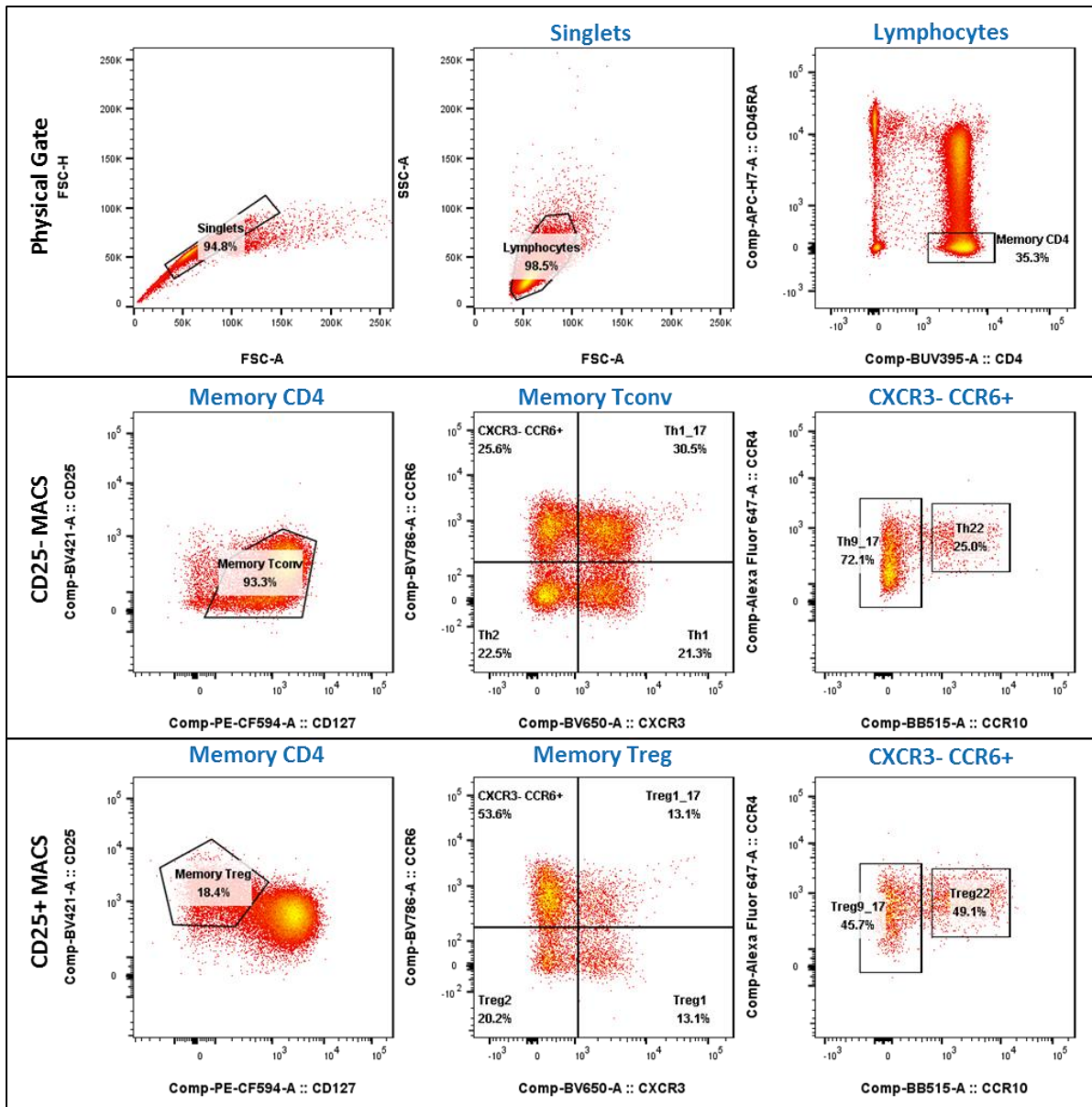


and hence were pooled together. To match the pooled Treg9/17, the Th9/17 were pooled as well during the sort.



**Figure 3.9: CD25 MACS purity FACS plot**

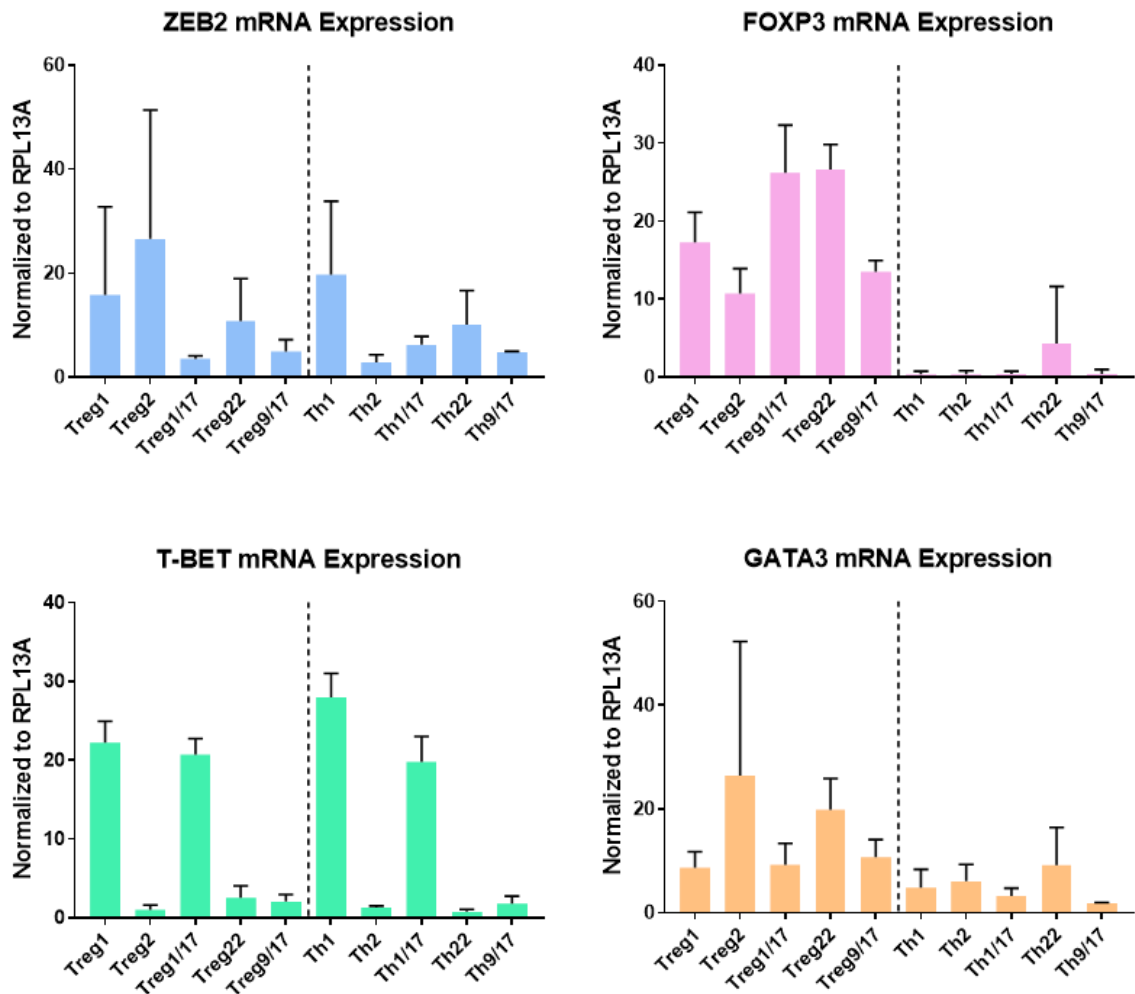
Flow cytometry plot showing purity of Treg and Tconv post CD25 MACS of enriched CD4+ T cells. Treg (CD25hi, CD127lo) and Tconv (CD25lo, CD127hi) were gated on both CD25-MACS (left) fraction and CD25+ MACS (right) fraction of the same sample. CD25+ MACS (left) fraction was only used for Tconv helper lineage population isolation whereas CD25-MACS (right) fraction was used for Treg helper lineage population isolation. Representative data from 1 donor, n=3 independent donors.



**Figure 3.10: Chemokine receptor helper lineages profiling and sorting strategy of post CD25 MACS Treg and Tconv**

Flow cytometry plot showing post RosetteSep™ enriched CD4 and post CD25 MACs. CD25- (enriched Tconv) and CD25+ (enriched Treg) MACS fractions were labelled separately with CD4, CD25, CD127, CD45RA, CXCR3, CCR6, CCR4, CCR10. Cells were first physically gated on singlets, lymphocytes and memory CD4+ (top). CD25- MACS fraction was gated for Tconv (middle) and CD25+ fraction was gated for Treg (bottom). Representative data from 1 donor, n=3 independent donors.

Using this approach, I found that ZEB2 mRNA expression was highly variable and was not increased in any of the lineages in either Tconv or in Treg and Tconv (Figure 3.11 below). FOXP3 mRNA expression was highly expressed across all Treg helper lineages compared with its expression in Tconv helper lineages. T-bet mRNA expression was highly expressed in Treg1, Treg1/17, Th1 and Th1/17 helper lineages. Surprisingly, I found that GATA3 expression was not significantly enriched in Th2 and Treg2. Since we know that ZEB2 is induced by T-bet (Dominguez et al., 2015, Omilusik et al., 2015, van Helden et al., 2015), I anticipated that ZEB2 expression would correlate with that of T-bet expression. Conversely, in Treg lineages where FOXP3 repression of ZEB2 was assumed, I anticipated low ZEB2 expression. The ZEB2 results presented here are inconclusive. However, further investigation suggested that these inconsistent findings might have been owing to the CD25 MACs enrichment step itself, whereby CD25<sup>hi</sup> cells in the gated CD25<sup>-</sup> MACS fraction of Tconv helper lineages were contaminating the CD25<sup>+</sup> MACS fraction (used for Treg isolation as observed in Figure 3.9), and CD25<sup>+</sup> Tconv were excluded from the CD127<sup>hi</sup> Tconv population. Hence, by performing a CD25 MACS pre-enrichment step before the chemokine receptor panel sort, purification may bias for Tconv cells that are CD25<sup>-</sup> and such bias may not identify high expressing ZEB2 helper lineages correctly, resulting in false positive ZEB2 in the overall Treg helper lineage population, and false negative ZEB2 expression in the Tconv population.



**Figure 3.11: ZEB2, T-bet and FOXP3 mRNA expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment**

ZEB2 (blue), FOXP3 (pink), T-bet (green) and GATA3 (orange) expression in helper lineage populations of Treg and Tconv. Relative abundance of ZEB2, FOXP3, T-bet and GATA3 were normalised to reference gene RPL13A and plotted with mean + SEM. As there are multiple comparisons, the statistical significance is presented in a separate table, to keep the figures clear. Statistics were carried out with ordinary one-way Anova multiple comparison Fisher's Least Significant Difference (LSD) test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$  as shown in Table 3.9, Table 3.10, Table 3.11 and Table 3.12,  $n = 3$  independent donors.

**Table 3.9: Anova Fisher's LSD test on ZEB2 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	-10.75	-29.48 to 7.986	ns	0.2446
<i>Treg1 vs. Treg1/17</i>	12.32	-4.436 to 29.07	ns	0.1403
<i>Treg1 vs. Treg22</i>	5.054	-11.70 to 21.81	ns	0.5354
<i>Treg1 vs. Treg9/17</i>	10.92	-5.834 to 27.67	ns	0.1885
<i>Treg1 vs. Th1</i>	-3.843	-20.60 to 12.91	ns	0.6366
<i>Treg1 vs. Th2</i>	12.97	-3.789 to 29.72	ns	0.1218
<i>Treg1 vs. Th1/17</i>	9.551	-7.203 to 26.31	ns	0.2475
<i>Treg1 vs. Th22</i>	5.740	-11.01 to 22.49	ns	0.4821
<i>Treg1 vs. Th9/17</i>	11.08	-5.676 to 27.83	ns	0.1824
<i>Treg2 vs. Treg1/17</i>	23.07	4.333 to 41.80	*	0.0185
<i>Treg2 vs. Treg22</i>	15.80	-2.932 to 34.53	ns	0.0936
<i>Treg2 vs. Treg9/17</i>	21.67	2.935 to 40.40	*	0.0257
<i>Treg2 vs. Th1</i>	6.903	-11.83 to 25.64	ns	0.4500
<i>Treg2 vs. Th2</i>	23.71	4.980 to 42.44	*	0.0158
<i>Treg2 vs. Th1/17</i>	20.30	1.566 to 39.03	*	0.0352
<i>Treg2 vs. Th22</i>	16.49	-2.245 to 35.22	ns	0.0811
<i>Treg2 vs. Th9/17</i>	21.83	3.093 to 40.56	*	0.0247
<i>Treg1/17 vs. Treg22</i>	-7.265	-24.02 to 9.489	ns	0.3755
<i>Treg1/17 vs. Treg9/17</i>	-1.399	-18.15 to 15.36	ns	0.8631
<i>Treg1/17 vs. Th1</i>	-16.16	-32.92 to 0.5921	ns	0.0578
<i>Treg1/17 vs. Th2</i>	0.6463	-16.11 to 17.40	ns	0.9365
<i>Treg1/17 vs. Th1/17</i>	-2.768	-19.52 to 13.99	ns	0.7333
<i>Treg1/17 vs. Th22</i>	-6.579	-23.33 to 10.18	ns	0.4214
<i>Treg1/17 vs. Th9/17</i>	-1.241	-18.00 to 15.51	ns	0.8785
<i>Treg22 vs. Treg9/17</i>	5.866	-10.89 to 22.62	ns	0.4726
<i>Treg22 vs. Th1</i>	-8.897	-25.65 to 7.857	ns	0.2802
<i>Treg22 vs. Th2</i>	7.912	-8.843 to 24.67	ns	0.3354
<i>Treg22 vs. Th1/17</i>	4.498	-12.26 to 21.25	ns	0.5808
<i>Treg22 vs. Th22</i>	0.6864	-16.07 to 17.44	ns	0.9326
<i>Treg22 vs. Th9/17</i>	6.025	-10.73 to 22.78	ns	0.4609
<i>Treg9/17 vs. Th1</i>	-14.76	-31.52 to 1.991	ns	0.0808
<i>Treg9/17 vs. Th2</i>	2.045	-14.71 to 18.80	ns	0.8011
<i>Treg9/17 vs. Th1/17</i>	-1.369	-18.12 to 15.39	ns	0.8660
<i>Treg9/17 vs. Th22</i>	-5.180	-21.93 to 11.57	ns	0.5253
<i>Treg9/17 vs. Th9/17</i>	0.1581	-16.60 to 16.91	ns	0.9844
<i>Th1 vs. Th2</i>	16.81	0.05418 to 33.56	*	0.0493
<i>Th1 vs. Th1/17</i>	13.39	-3.360 to 30.15	ns	0.1106
<i>Th1 vs. Th22</i>	9.584	-7.171 to 26.34	ns	0.2460
<i>Th1 vs. Th9/17</i>	14.92	-1.833 to 31.68	ns	0.0778
<i>Th2 vs. Th1/17</i>	-3.414	-20.17 to 13.34	ns	0.6746
<i>Th2 vs. Th22</i>	-7.225	-23.98 to 9.529	ns	0.3780
<i>Th2 vs. Th9/17</i>	-1.887	-18.64 to 14.87	ns	0.8162
<i>Th1/17 vs. Th22</i>	-3.811	-20.57 to 12.94	ns	0.6394
<i>Th1/17 vs. Th9/17</i>	1.527	-15.23 to 18.28	ns	0.8507
<i>Th22 vs. Th9/17</i>	5.338	-11.42 to 22.09	ns	0.5129

**Table 3.10: Anova Fisher's LSD test on FOXP3 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	6.544	-0.3736 to 13.46	ns	0.0624
<i>Treg1 vs. Treg1/17</i>	-8.934	-15.12 to -2.746	**	0.0070
<i>Treg1 vs. Treg22</i>	-9.317	-15.50 to -3.129	**	0.0053
<i>Treg1 vs. Treg9/17</i>	3.756	-2.432 to 9.943	ns	0.2193
<i>Treg1 vs. Th1</i>	16.89	10.71 to 23.08	****	<0.0001
<i>Treg1 vs. Th2</i>	16.91	10.72 to 23.10	****	<0.0001
<i>Treg1 vs. Th1/17</i>	16.84	10.65 to 23.03	****	<0.0001
<i>Treg1 vs. Th22</i>	12.96	6.774 to 19.15	***	0.0003
<i>Treg1 vs. Th9/17</i>	16.85	10.66 to 23.03	****	<0.0001
<i>Treg2 vs. Treg1/17</i>	-15.48	-22.40 to -8.560	***	0.0002
<i>Treg2 vs. Treg22</i>	-15.86	-22.78 to -8.943	***	0.0001
<i>Treg2 vs. Treg9/17</i>	-2.789	-9.706 to 4.129	ns	0.4093
<i>Treg2 vs. Th1</i>	10.35	3.433 to 17.27	**	0.0055
<i>Treg2 vs. Th2</i>	10.36	3.446 to 17.28	**	0.0054
<i>Treg2 vs. Th1/17</i>	10.30	3.380 to 17.22	**	0.0057
<i>Treg2 vs. Th22</i>	6.417	-0.5006 to 13.33	ns	0.0672
<i>Treg2 vs. Th9/17</i>	10.30	3.385 to 17.22	**	0.0057
<i>Treg1/17 vs. Treg22</i>	-0.3833	-6.571 to 5.804	ns	0.8982
<i>Treg1/17 vs. Treg9/17</i>	12.69	6.502 to 18.88	***	0.0004
<i>Treg1/17 vs. Th1</i>	25.83	19.64 to 32.02	****	<0.0001
<i>Treg1/17 vs. Th2</i>	25.84	19.65 to 32.03	****	<0.0001
<i>Treg1/17 vs. Th1/17</i>	25.78	19.59 to 31.96	****	<0.0001
<i>Treg1/17 vs. Th22</i>	21.89	15.71 to 28.08	****	<0.0001
<i>Treg1/17 vs. Th9/17</i>	25.78	19.59 to 31.97	****	<0.0001
<i>Treg22 vs. Treg9/17</i>	13.07	6.885 to 19.26	***	0.0003
<i>Treg22 vs. Th1</i>	26.21	20.02 to 32.40	****	<0.0001
<i>Treg22 vs. Th2</i>	26.23	20.04 to 32.41	****	<0.0001
<i>Treg22 vs. Th1/17</i>	26.16	19.97 to 32.35	****	<0.0001
<i>Treg22 vs. Th22</i>	22.28	16.09 to 28.47	****	<0.0001
<i>Treg22 vs. Th9/17</i>	26.16	19.98 to 32.35	****	<0.0001
<i>Treg9/17 vs. Th1</i>	13.14	6.952 to 19.33	***	0.0003
<i>Treg9/17 vs. Th2</i>	13.15	6.965 to 19.34	***	0.0003
<i>Treg9/17 vs. Th1/17</i>	13.09	6.899 to 19.27	***	0.0003
<i>Treg9/17 vs. Th22</i>	9.206	3.018 to 15.39	**	0.0057
<i>Treg9/17 vs. Th9/17</i>	13.09	6.904 to 19.28	***	0.0003
<i>Th1 vs. Th2</i>	0.01353	-6.174 to 6.201	ns	0.9964
<i>Th1 vs. Th1/17</i>	-0.05262	-6.240 to 6.135	ns	0.9860
<i>Th1 vs. Th22</i>	-3.933	-10.12 to 2.254	ns	0.1991
<i>Th1 vs. Th9/17</i>	-0.04758	-6.235 to 6.140	ns	0.9873
<i>Th2 vs. Th1/17</i>	-0.06616	-6.254 to 6.121	ns	0.9824
<i>Th2 vs. Th22</i>	-3.947	-10.13 to 2.240	ns	0.1976
<i>Th2 vs. Th9/17</i>	-0.06111	-6.249 to 6.126	ns	0.9837
<i>Th1/17 vs. Th22</i>	-3.881	-10.07 to 2.307	ns	0.2049
<i>Th1/17 vs. Th9/17</i>	0.005045	-6.182 to 6.192	ns	0.9987
<i>Th22 vs. Th9/17</i>	3.886	-2.302 to 10.07	ns	0.2043

**Table 3.11: Anova Fisher's LSD test on T-bet expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	21.19	17.97 to 24.41	****	<0.0001
<i>Treg1 vs. Treg1/17</i>	1.532	-1.688 to 4.752	ns	0.3328
<i>Treg1 vs. Treg22</i>	19.72	16.50 to 22.94	****	<0.0001
<i>Treg1 vs. Treg9/17</i>	20.16	16.94 to 23.38	****	<0.0001
<i>Treg1 vs. Th1</i>	-5.733	-8.953 to -2.514	**	0.0014
<i>Treg1 vs. Th2</i>	20.94	17.72 to 24.16	****	<0.0001
<i>Treg1 vs. Th1/17</i>	2.435	-0.7851 to 5.654	ns	0.1304
<i>Treg1 vs. Th22</i>	21.44	18.22 to 24.66	****	<0.0001
<i>Treg1 vs. Th9/17</i>	20.44	17.22 to 23.66	****	<0.0001
<i>Treg2 vs. Treg1/17</i>	-19.66	-22.88 to -16.44	****	<0.0001
<i>Treg2 vs. Treg22</i>	-1.466	-4.686 to 1.753	ns	0.3534
<i>Treg2 vs. Treg9/17</i>	-1.030	-4.250 to 2.190	ns	0.5122
<i>Treg2 vs. Th1</i>	-26.92	-30.14 to -23.70	****	<0.0001
<i>Treg2 vs. Th2</i>	-0.2458	-3.466 to 2.974	ns	0.8751
<i>Treg2 vs. Th1/17</i>	-18.75	-21.97 to -15.53	****	<0.0001
<i>Treg2 vs. Th22</i>	0.2530	-2.967 to 3.473	ns	0.8715
<i>Treg2 vs. Th9/17</i>	-0.7512	-3.971 to 2.469	ns	0.6318
<i>Treg1/17 vs. Treg22</i>	18.19	14.97 to 21.41	****	<0.0001
<i>Treg1/17 vs. Treg9/17</i>	18.63	15.41 to 21.85	****	<0.0001
<i>Treg1/17 vs. Th1</i>	-7.265	-10.49 to -4.046	**	0.0001
<i>Treg1/17 vs. Th2</i>	19.41	16.19 to 22.63	****	<0.0001
<i>Treg1/17 vs. Th1/17</i>	0.9028	-2.317 to 4.123	ns	0.5651
<i>Treg1/17 vs. Th22</i>	19.91	16.69 to 23.13	****	<0.0001
<i>Treg1/17 vs. Th9/17</i>	18.91	15.69 to 22.13	****	<0.0001
<i>Treg22 vs. Treg9/17</i>	0.4363	-2.783 to 3.656	ns	0.7803
<i>Treg22 vs. Th1</i>	-25.46	-28.68 to -22.24	****	<0.0001
<i>Treg22 vs. Th2</i>	1.221	-1.999 to 4.440	ns	0.4384
<i>Treg22 vs. Th1/17</i>	-17.29	-20.51 to -14.07	****	<0.0001
<i>Treg22 vs. Th22</i>	1.719	-1.500 to 4.939	ns	0.2785
<i>Treg22 vs. Th9/17</i>	0.7152	-2.505 to 3.935	ns	0.6481
<i>Treg9/17 vs. Th1</i>	-25.89	-29.11 to -22.67	****	<0.0001
<i>Treg9/17 vs. Th2</i>	0.7842	-2.436 to 4.004	ns	0.6170
<i>Treg9/17 vs. Th1/17</i>	-17.72	-20.94 to -14.50	****	<0.0001
<i>Treg9/17 vs. Th22</i>	1.283	-1.937 to 4.503	ns	0.4156
<i>Treg9/17 vs. Th9/17</i>	0.2789	-2.941 to 3.499	ns	0.8585
<i>Th1 vs. Th2</i>	26.68	23.46 to 29.90	****	<0.0001
<i>Th1 vs. Th1/17</i>	8.168	4.948 to 11.39	****	<0.0001
<i>Th1 vs. Th22</i>	27.18	23.96 to 30.40	****	<0.0001
<i>Th1 vs. Th9/17</i>	26.17	22.95 to 29.39	****	<0.0001
<i>Th2 vs. Th1/17</i>	-18.51	-21.73 to -15.29	****	<0.0001
<i>Th2 vs. Th22</i>	0.4988	-2.721 to 3.719	ns	0.7499
<i>Th2 vs. Th9/17</i>	-0.5054	-3.725 to 2.714	ns	0.7468
<i>Th1/17 vs. Th22</i>	19.01	15.79 to 22.23	****	<0.0001
<i>Th1/17 vs. Th9/17</i>	18.00	14.78 to 21.22	****	<0.0001
<i>Th22 vs. Th9/17</i>	-1.004	-4.224 to 2.216	ns	0.5227

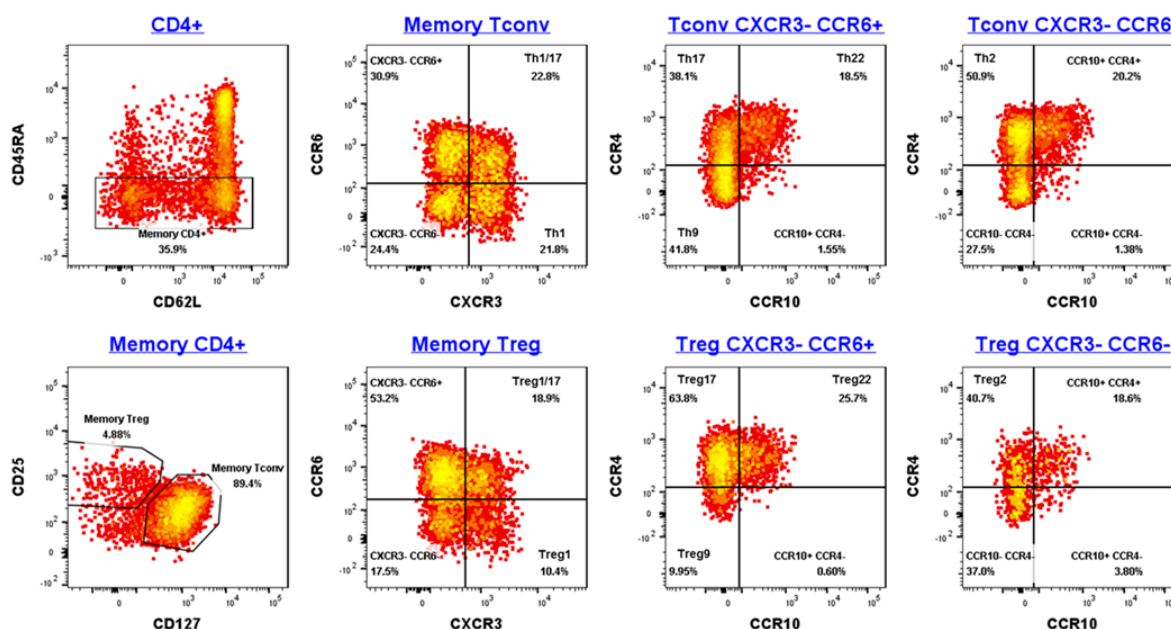
**Table 3.12: Anova Fisher's LSD test on GATA3 expression in helper lineage populations of Treg and Tconv with CD25 MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	-17.73	-37.92 to 2.458	ns	0.0788
<i>Treg1 vs. Treg1/17</i>	-0.5600	-20.75 to 19.63	ns	0.9519
<i>Treg1 vs. Treg22</i>	-11.15	-31.34 to 9.043	ns	0.2468
<i>Treg1 vs. Treg9/17</i>	-2.004	-22.19 to 18.19	ns	0.8294
<i>Treg1 vs. Th1</i>	3.861	-16.33 to 24.05	ns	0.6791
<i>Treg1 vs. Th2</i>	2.605	-17.58 to 22.79	ns	0.7796
<i>Treg1 vs. Th1/17</i>	5.431	-14.76 to 25.62	ns	0.5622
<i>Treg1 vs. Th22</i>	-0.5021	-20.69 to 19.69	ns	0.9569
<i>Treg1 vs. Th9/17</i>	6.807	-13.38 to 27.00	ns	0.4698
<i>Treg2 vs. Treg1/17</i>	17.17	-3.018 to 37.36	ns	0.0873
<i>Treg2 vs. Treg22</i>	6.585	-13.60 to 26.77	ns	0.4841
<i>Treg2 vs. Treg9/17</i>	15.73	-4.462 to 35.92	ns	0.1133
<i>Treg2 vs. Th1</i>	21.59	1.403 to 41.78	*	0.0384
<i>Treg2 vs. Th2</i>	20.34	0.1469 to 40.52	*	0.0486
<i>Treg2 vs. Th1/17</i>	23.16	2.973 to 43.35	*	0.0286
<i>Treg2 vs. Th22</i>	17.23	-2.960 to 37.42	ns	0.0864
<i>Treg2 vs. Th9/17</i>	24.54	4.349 to 44.73	*	0.0220
<i>Treg1/17 vs. Treg22</i>	-10.59	-30.78 to 9.602	ns	0.2698
<i>Treg1/17 vs. Treg9/17</i>	-1.444	-21.63 to 18.75	ns	0.8766
<i>Treg1/17 vs. Th1</i>	4.421	-15.77 to 24.61	ns	0.6361
<i>Treg1/17 vs. Th2</i>	3.165	-17.02 to 23.35	ns	0.7341
<i>Treg1/17 vs. Th1/17</i>	5.991	-14.20 to 26.18	ns	0.5234
<i>Treg1/17 vs. Th22</i>	0.05791	-20.13 to 20.25	ns	0.9950
<i>Treg1/17 vs. Th9/17</i>	7.367	-12.82 to 27.56	ns	0.4351
<i>Treg22 vs. Treg9/17</i>	9.143	-11.05 to 29.33	ns	0.3367
<i>Treg22 vs. Th1</i>	15.01	-5.182 to 35.20	ns	0.1287
<i>Treg22 vs. Th2</i>	13.75	-6.438 to 33.94	ns	0.1601
<i>Treg22 vs. Th1/17</i>	16.58	-3.612 to 36.77	ns	0.0972
<i>Treg22 vs. Th22</i>	10.64	-9.545 to 30.83	ns	0.2673
<i>Treg22 vs. Th9/17</i>	17.95	-2.235 to 38.14	ns	0.0757
<i>Treg9/17 vs. Th1</i>	5.865	-14.32 to 26.05	ns	0.5321
<i>Treg9/17 vs. Th2</i>	4.609	-15.58 to 24.80	ns	0.6221
<i>Treg9/17 vs. Th1/17</i>	7.435	-12.75 to 27.62	ns	0.4310
<i>Treg9/17 vs. Th22</i>	1.502	-18.69 to 21.69	ns	0.8717
<i>Treg9/17 vs. Th9/17</i>	8.811	-11.38 to 29.00	ns	0.3538
<i>Th1 vs. Th2</i>	-1.256	-21.45 to 18.93	ns	0.8925
<i>Th1 vs. Th1/17</i>	1.570	-18.62 to 21.76	ns	0.8659
<i>Th1 vs. Th22</i>	-4.363	-24.55 to 15.83	ns	0.6405
<i>Th1 vs. Th9/17</i>	2.946	-17.24 to 23.14	ns	0.7518
<i>Th2 vs. Th1/17</i>	2.826	-17.36 to 23.02	ns	0.7615
<i>Th2 vs. Th22</i>	-3.107	-23.30 to 17.08	ns	0.7388
<i>Th2 vs. Th9/17</i>	4.202	-15.99 to 24.39	ns	0.6527
<i>Th1/17 vs. Th22</i>	-5.933	-26.12 to 14.26	ns	0.5274
<i>Th1/17 vs. Th9/17</i>	1.376	-18.81 to 21.57	ns	0.8823
<i>Th22 vs. Th9/17</i>	7.309	-12.88 to 27.50	ns	0.4386



### 3.4.5 ZEB2 Expression in CD45RA MACS Enriched Purified Treg and Tconv Helper Lineages

The pre-enrichment step of Tconv and Treg using CD25 MACS prior to the chemokine panel sort purification led to an unexpected ZEB2 expression pattern across helper lineages, which was most likely due to contamination in the gated pools. Another approach is to use CD45RA MACS for pre-enrichment without affecting the CD25 status of the Tconv. As before, chemokine receptors CXCR3, CCR4, CCR6 and CCR10 (Table 3.3) were used to separate the helper lineages with revised Th17, Th9 and Th2 gating for both memory Treg and memory Tconv (Figure 3.12).



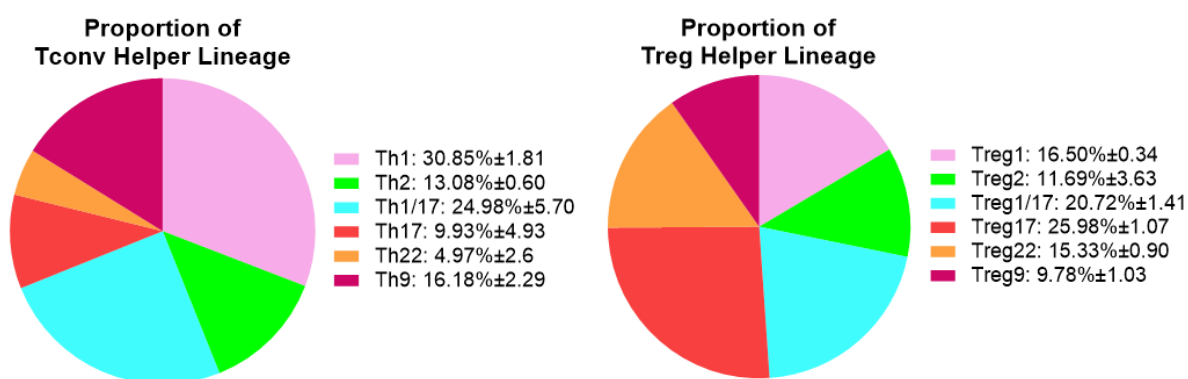
**Figure 3.12: Chemokine receptor helper lineages profiling and sorting strategy from post CD45RA- enriched MACS CD4+ T cell**

Flow cytometry plot showing post RosetteSep™ enriched CD4 and post CD45RA MACS. Enriched memory CD4 were labelled with surface antibodies from panel 3 including CD4, CD25, CD127, CD45RA, CXCR3, CCR6, CCR4, CCR10. Cells were first physically gated on singlets, lymphocytes, memory CD4+ and further divided into memory Treg and memory Tconv before separating into different helper lineages based on chemokine receptors expression. Sort and gating strategy as shown from 1 donor, n=3 independent donors.

#### Comparison of the helper lineages proportions in Treg and Tconv

First, I examined the proportion of the different lineage subsets in each of Tconv and Treg (Figure 3.13). I observed some interesting similarities and differences in the proportions of each

of the subsets between the Treg and Tconv populations. Whereas the Th1 population comprised the largest subset in Tconv (30.8%), in Treg this was not the largest subset (16.5% Treg1). In contrast, in Treg the Treg17 subset was the largest population (26%) but this population is relatively small in Tconv (9.9% Th17 in Tconv) subsets (Figure 3.13). Likewise, the Th9 population in Tconv (16.2% in Th9) was almost twice that of Treg (9.8% Treg9 in Treg) while the Treg22 population in Treg was over 3 times (15.3% Treg22 in Treg) that of the Th22 population in Tconv (5% Th22 in Tconv). The Treg/Th (1/17) and the Treg/Th (2) populations, however, were very similar in each of Treg and Tconv.



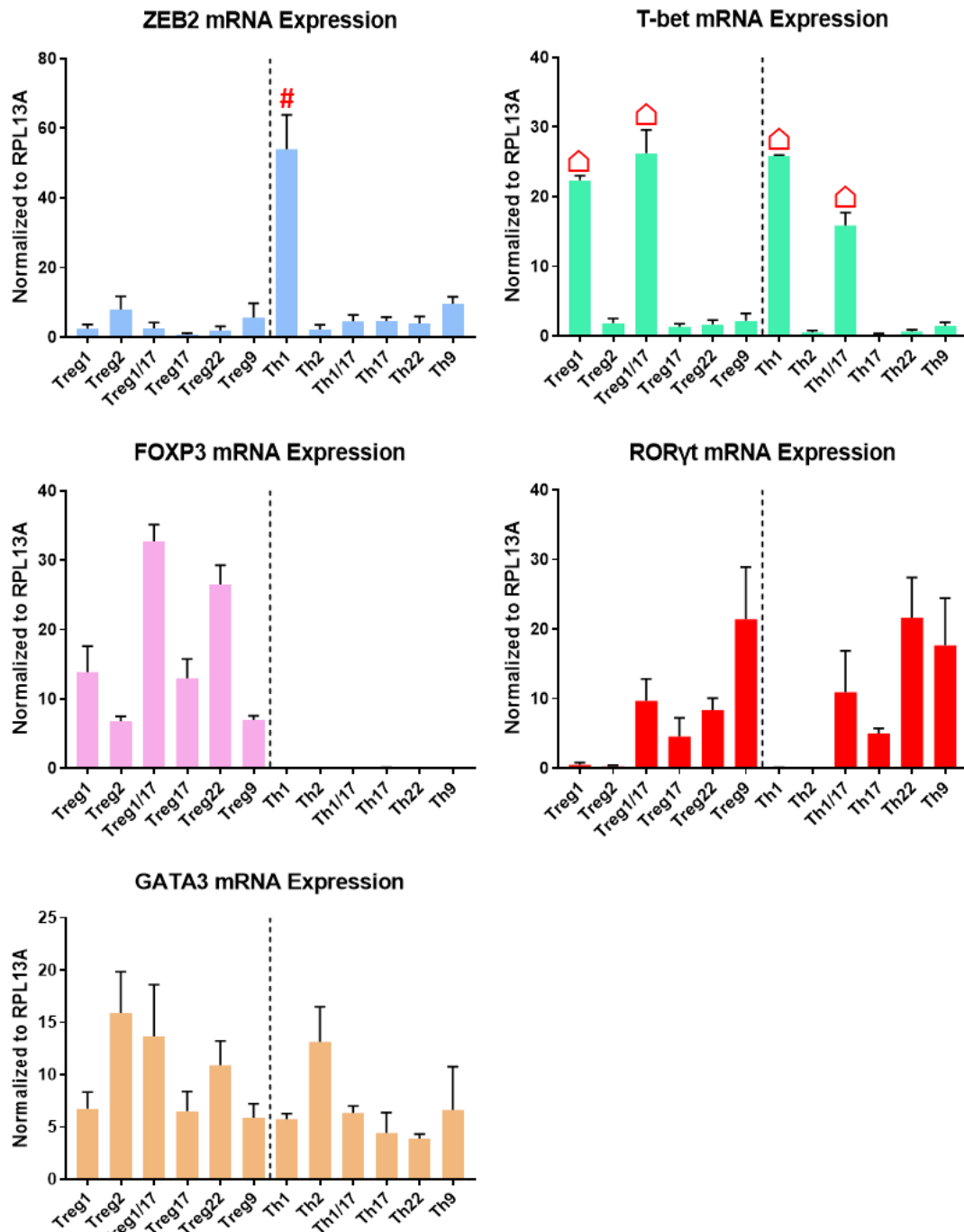
**Figure 3.13: Proportion of Tconv and Treg helper lineage populations**

Pie chart showing the proportion of Tconv (left) and Treg (right) helper lineage populations with SEM from 3 independent donors.

#### ZEB2 mRNA expression in Treg and Tconv helper lineage populations

To determine the expression pattern of ZEB2 mRNA in the Tconv and Treg lineage subsets, I used the gating strategy in Figure 3.12 to purify each subset. The sort gate was set more stringently to prevent “spill-over” between populations and ensure cells acquired are at the highest purity (>90%). I then isolated total RNA from each of the purified helper lineage populations in both Tconv and Treg. In order to ascertain the validity of this approach, I first confirmed that FOXP3 expression was highly specific to the Treg cells (in several subsets) and I then examined expression of ZEB2 mRNA across the lineage subsets. Interestingly, I found that ZEB2 is almost entirely confined to the Th1 cells (Figure 3.14). Since T-bet is known to

induce expression of ZEB2 in other immune cell types, I then compared the expression of the Th1 master transcription factor T-bet with that of ZEB2. Interestingly, although there was the highest T-bet expression in Th1 cells (as expected), T-bet had a broader expression profile (Figure 3.14). T-bet was also expressed in Th1/17 cells and in the Treg1 and Treg1/17 cell populations of Treg, indicating that expression of T-bet and ZEB2 do not necessarily correlate. As can be seen in Figure 3.14, expression of FOXP3 was high in several of the Treg subsets including Treg1/17 and Treg22 and was almost entirely absent in freshly isolated Tconv subsets. GATA3 expression was high in Tconv Th2 as expected and interestingly, although its expression was high in Treg Th2, it was also expressed in Treg Th1/17 and Treg Th22 subsets. ROR $\gamma$ t mRNA expression was enriched in CCR6<sup>+</sup> cells including Th1/17, Th17, Th22, Th9 of Treg and Tconv. Expression of the key lineage defining transcription factors for Th1(T-bet), Th2 (GATA3), Th17 (ROR $\gamma$ t) and Th1/17 (T-bet & ROR $\gamma$ t) confirmed the validity of this improved gating strategy compared to experiments performed previously in Section 3.4.4 in which CD25 MACs pre-enrichment were used (Figure 3.10).



**Figure 3.14: ZEB2, T-bet, RORγt and FOXP3 mRNA expression in helper lineage populations of Treg and Tconv**

ZEB2 (blue), T-bet (green), FOXP3 (pink), RORγt (red) and GATA3 (orange) expression in helper lineage populations of Treg and Tconv. Relative abundance of ZEB2, T-bet, FOXP3, RORγt and GATA3 were normalised to reference gene RPL13A and plotted with mean + SEM. As there are multiple comparisons, the statistical significance is presented in a separate table, to keep the figures clear. The annotation # indicates the level of ZEB2 in Th1 is significantly higher than other populations whereas populations annotated with Δ indicates the level of T-bet is not significantly different from one another. Statistics were carried out with ordinary one-way Anova multiple comparison Fisher's Least Significant Difference (LSD) test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$  as shown in Table 3.13, Table 3.14, Table 3.15, Table 3.16 and Table 3.17 below,  $n = 3$  independent donors, experiment carried out in triplicate.

**Table 3.13: Anova Fisher's LSD test on ZEB2 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	-5.501	-15.76 to 4.760	ns	0.2795
<i>Treg1 vs. Treg1/17</i>	-0.07827	-10.34 to 10.18	ns	0.9876
<i>Treg1 vs. Treg17</i>	1.660	-8.600 to 11.92	ns	0.7413
<i>Treg1 vs. Treg22</i>	0.4775	-9.783 to 10.74	ns	0.9243
<i>Treg1 vs. Treg9</i>	-3.190	-13.45 to 7.070	ns	0.5272
<i>Treg1 vs. Th1</i>	-51.59	-61.85 to -41.33	****	<0.0001
<i>Treg1 vs. Th2</i>	0.2523	-10.01 to 10.51	ns	0.9599
<i>Treg1 vs. Th1/17</i>	-2.117	-12.38 to 8.144	ns	0.6740
<i>Treg1 vs. Th17</i>	-2.237	-12.50 to 8.023	ns	0.6567
<i>Treg1 vs. Th22</i>	-1.493	-11.75 to 8.767	ns	0.7665
<i>Treg1 vs. Th9</i>	-7.144	-17.40 to 3.116	ns	0.1636
<i>Treg2 vs. Treg1/17</i>	5.422	-4.838 to 15.68	ns	0.2862
<i>Treg2 vs. Treg17</i>	7.161	-3.100 to 17.42	ns	0.1627
<i>Treg2 vs. Treg22</i>	5.978	-4.282 to 16.24	ns	0.2409
<i>Treg2 vs. Treg9</i>	2.311	-7.950 to 12.57	ns	0.6463
<i>Treg2 vs. Th1</i>	-46.09	-56.35 to -35.83	****	<0.0001
<i>Treg2 vs. Th2</i>	5.753	-4.508 to 16.01	ns	0.2586
<i>Treg2 vs. Th1/17</i>	3.384	-6.877 to 13.64	ns	0.5026
<i>Treg2 vs. Th17</i>	3.264	-6.997 to 13.52	ns	0.5178
<i>Treg2 vs. Th22</i>	4.007	-6.253 to 14.27	ns	0.4281
<i>Treg2 vs. Th9</i>	-1.643	-11.90 to 8.617	ns	0.7439
<i>Treg1/17 vs. Treg17</i>	1.739	-8.522 to 12.00	ns	0.7296
<i>Treg1/17 vs. Treg22</i>	0.5557	-9.705 to 10.82	ns	0.9119
<i>Treg1/17 vs. Treg9</i>	-3.112	-13.37 to 7.149	ns	0.5373
<i>Treg1/17 vs. Th1</i>	-51.51	-61.77 to -41.25	****	<0.0001
<i>Treg1/17 vs. Th2</i>	0.3306	-9.930 to 10.59	ns	0.9475
<i>Treg1/17 vs. Th1/17</i>	-2.039	-12.30 to 8.222	ns	0.6854
<i>Treg1/17 vs. Th17</i>	-2.159	-12.42 to 8.102	ns	0.6680
<i>Treg1/17 vs. Th22</i>	-1.415	-11.68 to 8.845	ns	0.7783
<i>Treg1/17 vs. Th9</i>	-7.066	-17.33 to 3.195	ns	0.1681
<i>Treg17 vs. Treg22</i>	-1.183	-11.44 to 9.078	ns	0.8140
<i>Treg17 vs. Treg9</i>	-4.850	-15.11 to 5.410	ns	0.3390
<i>Treg17 vs. Th1</i>	-53.25	-63.51 to -42.99	****	<0.0001
<i>Treg17 vs. Th2</i>	-1.408	-11.67 to 8.853	ns	0.7794
<i>Treg17 vs. Th1/17</i>	-3.777	-14.04 to 6.483	ns	0.4548
<i>Treg17 vs. Th17</i>	-3.897	-14.16 to 6.363	ns	0.4407
<i>Treg17 vs. Th22</i>	-3.154	-13.41 to 7.107	ns	0.5318
<i>Treg17 vs. Th9</i>	-8.804	-19.06 to 1.456	ns	0.0893
<i>Treg22 vs. Treg9</i>	-3.667	-13.93 to 6.593	ns	0.4678
<i>Treg22 vs. Th1</i>	-52.06	-62.32 to -41.80	****	<0.0001
<i>Treg22 vs. Th2</i>	-0.2252	-10.49 to 10.04	ns	0.9642
<i>Treg22 vs. Th1/17</i>	-2.594	-12.85 to 7.666	ns	0.6066
<i>Treg22 vs. Th17</i>	-2.715	-12.98 to 7.546	ns	0.5901
<i>Treg22 vs. Th22</i>	-1.971	-12.23 to 8.290	ns	0.6953
<i>Treg22 vs. Th9</i>	-7.621	-17.88 to 2.639	ns	0.1383
<i>Treg9 vs. Th1</i>	-48.40	-58.66 to -38.14	****	<0.0001
<i>Treg9 vs. Th2</i>	3.442	-6.818 to 13.70	ns	0.4953
<i>Treg9 vs. Th1/17</i>	1.073	-9.187 to 11.33	ns	0.8309
<i>Treg9 vs. Th17</i>	0.9529	-9.308 to 11.21	ns	0.8496
<i>Treg9 vs. Th22</i>	1.697	-8.564 to 11.96	ns	0.7359
<i>Treg9 vs. Th9</i>	-3.954	-14.21 to 6.306	ns	0.4342
<i>Th1 vs. Th2</i>	51.84	41.58 to 62.10	****	<0.0001
<i>Th1 vs. Th1/17</i>	49.47	39.21 to 59.73	****	<0.0001
<i>Th1 vs. Th17</i>	49.35	39.09 to 59.61	****	<0.0001
<i>Th1 vs. Th22</i>	50.09	39.83 to 60.35	****	<0.0001
<i>Th1 vs. Th9</i>	44.44	34.18 to 54.70	****	<0.0001
<i>Th2 vs. Th1/17</i>	-2.369	-12.63 to 7.891	ns	0.6380

<i>Th2 vs. Th17</i>	-2.489	-12.75 to 7.771	ns	0.6211
<i>Th2 vs. Th22</i>	-1.746	-12.01 to 8.515	ns	0.7285
<i>Th2 vs. Th9</i>	-7.396	-17.66 to 2.864	ns	0.1498
<i>Th1/17 vs. Th17</i>	-0.1202	-10.38 to 10.14	ns	0.9809
<i>Th1/17 vs. Th22</i>	0.6235	-9.637 to 10.88	ns	0.9012
<i>Th1/17 vs. Th9</i>	-5.027	-15.29 to 5.233	ns	0.3220
<i>Th17 vs. Th22</i>	0.7437	-9.517 to 11.00	ns	0.8823
<i>Th17 vs. Th9</i>	-4.907	-15.17 to 5.354	ns	0.3335
<i>Th22 vs. Th9</i>	-5.651	-15.91 to 4.610	ns	0.2669

**Table 3.14: Anova Fisher's LSD test on T-bet expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	20.53	16.98 to 24.08	****	<0.0001
<i>Treg1 vs. Treg1/17</i>	-3.903	-7.450 to -0.3566	*	0.0324
<i>Treg1 vs. Treg17</i>	21.02	17.47 to 24.57	****	<0.0001
<i>Treg1 vs. Treg22</i>	20.68	17.14 to 24.23	****	<0.0001
<i>Treg1 vs. Treg9</i>	20.17	16.62 to 23.71	****	<0.0001
<i>Treg1 vs. Th1</i>	-3.480	-7.027 to 0.06627	ns	0.0541
<i>Treg1 vs. Th2</i>	21.83	18.28 to 25.38	****	<0.0001
<i>Treg1 vs. Th1/17</i>	6.476	2.929 to 10.02	***	0.0009
<i>Treg1 vs. Th17</i>	22.09	18.55 to 25.64	****	<0.0001
<i>Treg1 vs. Th22</i>	21.71	18.17 to 25.26	****	<0.0001
<i>Treg1 vs. Th9</i>	20.92	17.37 to 24.46	****	<0.0001
<i>Treg2 vs. Treg1/17</i>	-24.43	-27.98 to -20.89	****	<0.0001
<i>Treg2 vs. Treg17</i>	0.4919	-3.054 to 4.038	ns	0.7771
<i>Treg2 vs. Treg22</i>	0.1540	-3.392 to 3.700	ns	0.9293
<i>Treg2 vs. Treg9</i>	-0.3620	-3.908 to 3.184	ns	0.8349
<i>Treg2 vs. Th1</i>	-24.01	-27.56 to -20.46	****	<0.0001
<i>Treg2 vs. Th2</i>	1.299	-2.247 to 4.846	ns	0.4569
<i>Treg2 vs. Th1/17</i>	-14.05	-17.60 to -10.51	****	<0.0001
<i>Treg2 vs. Th17</i>	1.565	-1.981 to 5.112	ns	0.3714
<i>Treg2 vs. Th22</i>	1.185	-2.362 to 4.731	ns	0.4971
<i>Treg2 vs. Th9</i>	0.3868	-3.160 to 3.933	ns	0.8238
<i>Treg1/17 vs. Treg17</i>	24.92	21.38 to 28.47	****	<0.0001
<i>Treg1/17 vs. Treg22</i>	24.59	21.04 to 28.13	****	<0.0001
<i>Treg1/17 vs. Treg9</i>	24.07	20.52 to 27.62	****	<0.0001
<i>Treg1/17 vs. Th1</i>	0.4229	-3.124 to 3.969	ns	0.8077
<i>Treg1/17 vs. Th2</i>	25.73	22.19 to 29.28	****	<0.0001
<i>Treg1/17 vs. Th1/17</i>	10.38	6.832 to 13.93	****	<0.0001
<i>Treg1/17 vs. Th17</i>	26.00	22.45 to 29.54	****	<0.0001
<i>Treg1/17 vs. Th22</i>	25.62	22.07 to 29.16	****	<0.0001
<i>Treg1/17 vs. Th9</i>	24.82	21.27 to 28.37	****	<0.0001
<i>Treg17 vs. Treg22</i>	-0.3380	-3.884 to 3.208	ns	0.8457
<i>Treg17 vs. Treg9</i>	-0.8539	-4.400 to 2.693	ns	0.6237
<i>Treg17 vs. Th1</i>	-24.50	-28.05 to -20.96	****	<0.0001
<i>Treg17 vs. Th2</i>	0.8074	-2.739 to 4.354	ns	0.6427
<i>Treg17 vs. Th1/17</i>	-14.55	-18.09 to -11.00	****	<0.0001
<i>Treg17 vs. Th17</i>	1.073	-2.473 to 4.620	ns	0.5381
<i>Treg17 vs. Th22</i>	0.6928	-2.854 to 4.239	ns	0.6904
<i>Treg17 vs. Th9</i>	-0.1051	-3.652 to 3.441	ns	0.9517
<i>Treg22 vs. Treg9</i>	-0.5160	-4.062 to 3.030	ns	0.7666
<i>Treg22 vs. Th1</i>	-24.16	-27.71 to -20.62	****	<0.0001
<i>Treg22 vs. Th2</i>	1.145	-2.401 to 4.692	ns	0.5114
<i>Treg22 vs. Th1/17</i>	-14.21	-17.75 to -10.66	****	<0.0001
<i>Treg22 vs. Th17</i>	1.411	-2.135 to 4.958	ns	0.4196
<i>Treg22 vs. Th22</i>	1.031	-2.516 to 4.577	ns	0.5542
<i>Treg22 vs. Th9</i>	0.2328	-3.314 to 3.779	ns	0.8933
<i>Treg9 vs. Th1</i>	-23.65	-27.19 to -20.10	****	<0.0001
<i>Treg9 vs. Th2</i>	1.661	-1.885 to 5.208	ns	0.3433
<i>Treg9 vs. Th1/17</i>	-13.69	-17.24 to -10.15	****	<0.0001
<i>Treg9 vs. Th17</i>	1.927	-1.619 to 5.474	ns	0.2731
<i>Treg9 vs. Th22</i>	1.547	-2.000 to 5.093	ns	0.3770
<i>Treg9 vs. Th9</i>	0.7488	-2.798 to 4.295	ns	0.6669
<i>Th1 vs. Th2</i>	25.31	21.76 to 28.86	****	<0.0001
<i>Th1 vs. Th1/17</i>	9.956	6.409 to 13.50	****	<0.0001
<i>Th1 vs. Th17</i>	25.57	22.03 to 29.12	****	<0.0001
<i>Th1 vs. Th22</i>	25.19	21.65 to 28.74	****	<0.0001
<i>Th1 vs. Th9</i>	24.40	20.85 to 27.94	****	<0.0001
<i>Th2 vs. Th1/17</i>	-15.35	-18.90 to -11.81	****	<0.0001

<i>Th2 vs. Th17</i>	0.2659	-3.281 to 3.812	ns	0.8783
<i>Th2 vs. Th22</i>	-0.1146	-3.661 to 3.432	ns	0.9474
<i>Th2 vs. Th9</i>	-0.9125	-4.459 to 2.634	ns	0.6003
<i>Th1/17 vs. Th17</i>	15.62	12.07 to 19.17	****	<0.0001
<i>Th1/17 vs. Th22</i>	15.24	11.69 to 18.78	****	<0.0001
<i>Th1/17 vs. Th9</i>	14.44	10.89 to 17.99	****	<0.0001
<i>Th17 vs. Th22</i>	-0.3804	-3.927 to 3.166	ns	0.8267
<i>Th17 vs. Th9</i>	-1.178	-4.725 to 2.368	ns	0.4994
<i>Th22 vs. Th9</i>	-0.7979	-4.344 to 2.748	ns	0.6466



**Table 3.15: Anova Fisher's LSD test on FOXP3 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	7.069	1.998 to 12.14	**	0.0083
<i>Treg1 vs. Treg1/17</i>	-18.85	-23.93 to -13.78	****	<0.0001
<i>Treg1 vs. Treg17</i>	0.9219	-4.150 to 5.994	ns	0.7109
<i>Treg1 vs. Treg22</i>	-12.62	-17.69 to -7.548	****	<0.0001
<i>Treg1 vs. Treg9</i>	6.901	1.829 to 11.97	**	0.0097
<i>Treg1 vs. Th1</i>	13.83	8.756 to 18.90	****	<0.0001
<i>Treg1 vs. Th2</i>	13.83	8.758 to 18.90	****	<0.0001
<i>Treg1 vs. Th1/17</i>	13.80	8.724 to 18.87	****	<0.0001
<i>Treg1 vs. Th17</i>	13.77	8.699 to 18.84	****	<0.0001
<i>Treg1 vs. Th22</i>	13.83	8.763 to 18.91	****	<0.0001
<i>Treg1 vs. Th9</i>	13.79	8.723 to 18.87	****	<0.0001
<i>Treg2 vs. Treg1/17</i>	-25.92	-30.99 to -20.85	****	<0.0001
<i>Treg2 vs. Treg17</i>	-6.148	-11.22 to -1.076	*	0.0196
<i>Treg2 vs. Treg22</i>	-19.69	-24.76 to -14.62	****	<0.0001
<i>Treg2 vs. Treg9</i>	-0.1681	-5.240 to 4.904	ns	0.9460
<i>Treg2 vs. Th1</i>	6.759	1.687 to 11.83	*	0.0111
<i>Treg2 vs. Th2</i>	6.760	1.689 to 11.83	*	0.0111
<i>Treg2 vs. Th1/17</i>	6.726	1.654 to 11.80	*	0.0115
<i>Treg2 vs. Th17</i>	6.702	1.630 to 11.77	*	0.0117
<i>Treg2 vs. Th22</i>	6.765	1.693 to 11.84	*	0.0111
<i>Treg2 vs. Th9</i>	6.725	1.653 to 11.80	*	0.0115
<i>Treg1/17 vs. Treg17</i>	19.78	14.70 to 24.85	****	<0.0001
<i>Treg1/17 vs. Treg22</i>	6.234	1.162 to 11.31	*	0.0181
<i>Treg1/17 vs. Treg9</i>	25.75	20.68 to 30.83	****	<0.0001
<i>Treg1/17 vs. Th1</i>	32.68	27.61 to 37.75	****	<0.0001
<i>Treg1/17 vs. Th2</i>	32.68	27.61 to 37.76	****	<0.0001
<i>Treg1/17 vs. Th1/17</i>	32.65	27.58 to 37.72	****	<0.0001
<i>Treg1/17 vs. Th17</i>	32.62	27.55 to 37.70	****	<0.0001
<i>Treg1/17 vs. Th22</i>	32.69	27.62 to 37.76	****	<0.0001
<i>Treg1/17 vs. Th9</i>	32.65	27.58 to 37.72	****	<0.0001
<i>Treg17 vs. Treg22</i>	-13.54	-18.61 to -8.470	****	<0.0001
<i>Treg17 vs. Treg9</i>	5.979	0.9076 to 11.05	*	0.0228
<i>Treg17 vs. Th1</i>	12.91	7.834 to 17.98	****	<0.0001
<i>Treg17 vs. Th2</i>	12.91	7.836 to 17.98	****	<0.0001
<i>Treg17 vs. Th1/17</i>	12.87	7.802 to 17.95	****	<0.0001
<i>Treg17 vs. Th17</i>	12.85	7.778 to 17.92	****	<0.0001
<i>Treg17 vs. Th22</i>	12.91	7.841 to 17.98	****	<0.0001
<i>Treg17 vs. Th9</i>	12.87	7.801 to 17.94	****	<0.0001
<i>Treg22 vs. Treg9</i>	19.52	14.45 to 24.59	****	<0.0001
<i>Treg22 vs. Th1</i>	26.45	21.38 to 31.52	****	<0.0001
<i>Treg22 vs. Th2</i>	26.45	21.38 to 31.52	****	<0.0001
<i>Treg22 vs. Th1/17</i>	26.42	21.34 to 31.49	****	<0.0001
<i>Treg22 vs. Th17</i>	26.39	21.32 to 31.46	****	<0.0001
<i>Treg22 vs. Th22</i>	26.45	21.38 to 31.53	****	<0.0001
<i>Treg22 vs. Th9</i>	26.41	21.34 to 31.49	****	<0.0001
<i>Treg9 vs. Th1</i>	6.927	1.855 to 12.00	**	0.0095
<i>Treg9 vs. Th2</i>	6.928	1.857 to 12.00	**	0.0095
<i>Treg9 vs. Th1/17</i>	6.894	1.822 to 11.97	**	0.0098
<i>Treg9 vs. Th17</i>	6.870	1.798 to 11.94	*	0.0100
<i>Treg9 vs. Th22</i>	6.933	1.861 to 12.01	**	0.0094
<i>Treg9 vs. Th9</i>	6.893	1.821 to 11.96	**	0.0098
<i>Th1 vs. Th2</i>	0.001734	-5.070 to 5.074	ns	0.9994
<i>Th1 vs. Th1/17</i>	-0.03249	-5.104 to 5.039	ns	0.9896
<i>Th1 vs. Th17</i>	-0.05680	-5.129 to 5.015	ns	0.9818
<i>Th1 vs. Th22</i>	0.006435	-5.065 to 5.078	ns	0.9979
<i>Th1 vs. Th9</i>	-0.03374	-5.106 to 5.038	ns	0.9892
<i>Th2 vs. Th1/17</i>	-0.03422	-5.106 to 5.038	ns	0.9890

<i>Th2 vs. Th17</i>	-0.05853	-5.130 to 5.013	ns	0.9812
<i>Th2 vs. Th22</i>	0.004701	-5.067 to 5.077	ns	0.9985
<i>Th2 vs. Th9</i>	-0.03547	-5.107 to 5.036	ns	0.9886
<i>Th1/17 vs. Th17</i>	-0.02431	-5.096 to 5.048	ns	0.9922
<i>Th1/17 vs. Th22</i>	0.03893	-5.033 to 5.111	ns	0.9875
<i>Th1/17 vs. Th9</i>	-0.001248	-5.073 to 5.071	ns	0.9996
<i>Th17 vs. Th22</i>	0.06323	-5.009 to 5.135	ns	0.9797
<i>Th17 vs. Th9</i>	0.02306	-5.049 to 5.095	ns	0.9926
<i>Th22 vs. Th9</i>	-0.04017	-5.112 to 5.032	ns	0.9871

**Table 3.16: Anova Fisher's LSD test on ROR $\gamma$ t expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	0.2152	-11.43 to 11.86	ns	0.9699
<i>Treg1 vs. Treg1/17</i>	-9.197	-20.84 to 2.448	ns	0.1162
<i>Treg1 vs. Treg17</i>	-4.092	-15.74 to 7.553	ns	0.4753
<i>Treg1 vs. Treg22</i>	-7.900	-19.55 to 3.745	ns	0.1742
<i>Treg1 vs. Treg9</i>	-20.95	-32.60 to -9.307	**	0.0011
<i>Treg1 vs. Th1</i>	0.3758	-11.27 to 12.02	ns	0.9474
<i>Treg1 vs. Th2</i>	0.4545	-11.19 to 12.10	ns	0.9365
<i>Treg1 vs. Th1/17</i>	-10.45	-22.10 to 1.194	ns	0.0763
<i>Treg1 vs. Th17</i>	-4.497	-16.14 to 7.148	ns	0.4333
<i>Treg1 vs. Th22</i>	-21.16	-32.81 to -9.516	***	0.0010
<i>Treg1 vs. Th9</i>	-17.19	-28.84 to -5.549	**	0.0055
<i>Treg2 vs. Treg1/17</i>	-9.412	-21.06 to 2.233	ns	0.1083
<i>Treg2 vs. Treg17</i>	-4.307	-15.95 to 7.338	ns	0.4527
<i>Treg2 vs. Treg22</i>	-8.116	-19.76 to 3.529	ns	0.1632
<i>Treg2 vs. Treg9</i>	-21.17	-32.81 to -9.522	***	0.0010
<i>Treg2 vs. Th1</i>	0.1606	-11.48 to 11.81	ns	0.9775
<i>Treg2 vs. Th2</i>	0.2392	-11.41 to 11.88	ns	0.9665
<i>Treg2 vs. Th1/17</i>	-10.67	-22.31 to 0.9788	ns	0.0708
<i>Treg2 vs. Th17</i>	-4.712	-16.36 to 6.933	ns	0.4119
<i>Treg2 vs. Th22</i>	-21.38	-33.02 to -9.731	***	0.0009
<i>Treg2 vs. Th9</i>	-17.41	-29.05 to -5.764	**	0.0051
<i>Treg1/17 vs. Treg17</i>	5.105	-6.540 to 16.75	ns	0.3746
<i>Treg1/17 vs. Treg22</i>	1.296	-10.35 to 12.94	ns	0.8202
<i>Treg1/17 vs. Treg9</i>	-11.76	-23.40 to -0.1100	*	0.0480
<i>Treg1/17 vs. Th1</i>	9.573	-2.073 to 21.22	ns	0.1027
<i>Treg1/17 vs. Th2</i>	9.651	-1.994 to 21.30	ns	0.1001
<i>Treg1/17 vs. Th1/17</i>	-1.254	-12.90 to 10.39	ns	0.8259
<i>Treg1/17 vs. Th17</i>	4.700	-6.945 to 16.34	ns	0.4131
<i>Treg1/17 vs. Th22</i>	-11.96	-23.61 to -0.3191	*	0.0445
<i>Treg1/17 vs. Th9</i>	-7.998	-19.64 to 3.648	ns	0.1692
<i>Treg17 vs. Treg22</i>	-3.808	-15.45 to 7.837	ns	0.5061
<i>Treg17 vs. Treg9</i>	-16.86	-28.50 to -5.215	**	0.0064
<i>Treg17 vs. Th1</i>	4.468	-7.177 to 16.11	ns	0.4362
<i>Treg17 vs. Th2</i>	4.546	-7.099 to 16.19	ns	0.4283
<i>Treg17 vs. Th1/17</i>	-6.359	-18.00 to 5.286	ns	0.2709
<i>Treg17 vs. Th17</i>	-0.4049	-12.05 to 11.24	ns	0.9434
<i>Treg17 vs. Th22</i>	-17.07	-28.71 to -5.424	**	0.0058
<i>Treg17 vs. Th9</i>	-13.10	-24.75 to -1.457	*	0.0290
<i>Treg22 vs. Treg9</i>	-13.05	-24.70 to -1.406	*	0.0296
<i>Treg22 vs. Th1</i>	8.276	-3.369 to 19.92	ns	0.1554
<i>Treg22 vs. Th2</i>	8.355	-3.290 to 20.00	ns	0.1517
<i>Treg22 vs. Th1/17</i>	-2.551	-14.20 to 9.094	ns	0.6553
<i>Treg22 vs. Th17</i>	3.403	-8.242 to 15.05	ns	0.5520
<i>Treg22 vs. Th22</i>	-13.26	-24.91 to -1.615	*	0.0273
<i>Treg22 vs. Th9</i>	-9.294	-20.94 to 2.351	ns	0.1126
<i>Treg9 vs. Th1</i>	21.33	9.682 to 32.97	***	0.0009
<i>Treg9 vs. Th2</i>	21.41	9.761 to 33.05	***	0.0009
<i>Treg9 vs. Th1/17</i>	10.50	-1.144 to 22.15	ns	0.0750
<i>Treg9 vs. Th17</i>	16.45	4.810 to 28.10	**	0.0076
<i>Treg9 vs. Th22</i>	-0.2091	-11.85 to 11.44	ns	0.9707
<i>Treg9 vs. Th9</i>	3.758	-7.888 to 15.40	ns	0.5118
<i>Th1 vs. Th2</i>	0.07863	-11.57 to 11.72	ns	0.9890
<i>Th1 vs. Th1/17</i>	-10.83	-22.47 to 0.8182	ns	0.0670
<i>Th1 vs. Th17</i>	-4.873	-16.52 to 6.772	ns	0.3963
<i>Th1 vs. Th22</i>	-21.54	-33.18 to -9.892	***	0.0008
<i>Th1 vs. Th9</i>	-17.57	-29.22 to -5.925	**	0.0047
<i>Th2 vs. Th1/17</i>	-10.91	-22.55 to 0.7396	ns	0.0651

<i>Th2 vs. Th17</i>	-4.951	-16.60 to 6.694	ns	0.3889
<i>Th2 vs. Th22</i>	-21.62	-33.26 to -9.970	***	0.0008
<i>Th2 vs. Th9</i>	-17.65	-29.29 to -6.004	**	0.0046
<i>Th1/17 vs. Th17</i>	5.954	-5.691 to 17.60	ns	0.3018
<i>Th1/17 vs. Th22</i>	-10.71	-22.35 to 0.9353	ns	0.0698
<i>Th1/17 vs. Th9</i>	-6.743	-18.39 to 4.902	ns	0.2437
<i>Th17 vs. Th22</i>	-16.66	-28.31 to -5.019	**	0.0069
<i>Th17 vs. Th9</i>	-12.70	-24.34 to -1.052	*	0.0339
<i>Th22 vs. Th9</i>	3.967	-7.678 to 15.61	ns	0.4888

**Table 3.17: Anova Fisher's LSD test on GATA3 expression in helper lineage populations of Treg and Tconv with CD45RA MACS enrichment**

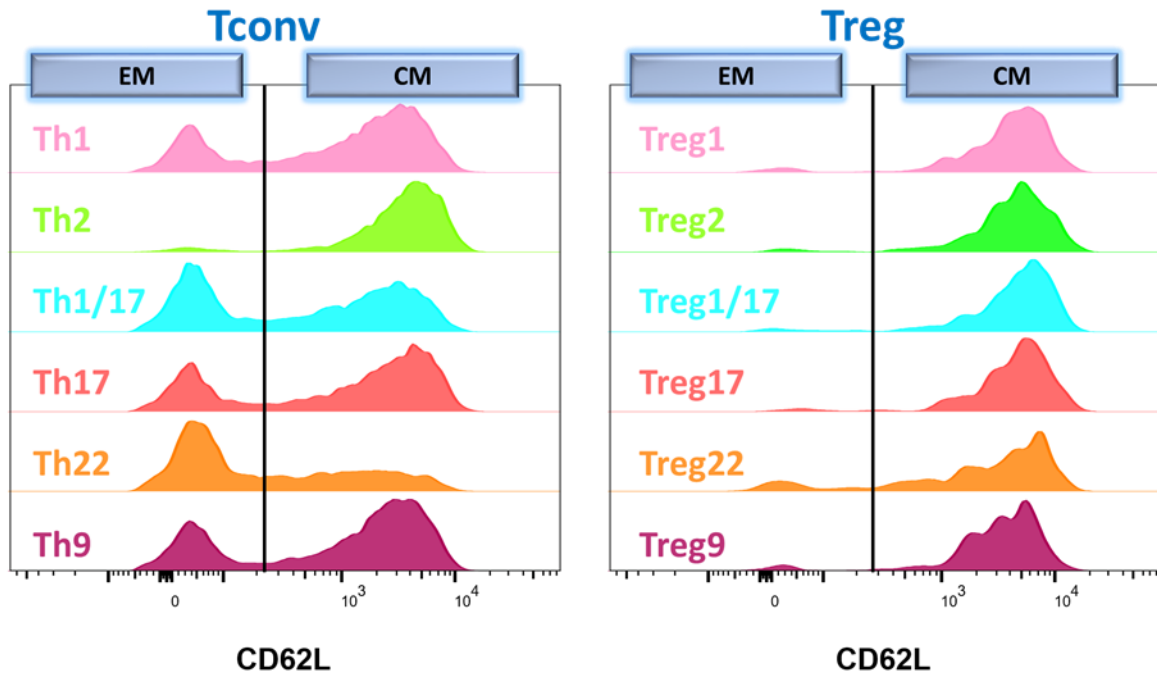
Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Treg1 vs. Treg2</i>	-9.178	-16.99 to -1.369	*	0.0232
<i>Treg1 vs. Treg1/17</i>	-6.935	-14.74 to 0.8745	ns	0.0793
<i>Treg1 vs. Treg17</i>	0.2209	-7.588 to 8.030	ns	0.9539
<i>Treg1 vs. Treg22</i>	-4.152	-11.96 to 3.657	ns	0.2834
<i>Treg1 vs. Treg9</i>	0.8413	-6.968 to 8.650	ns	0.8259
<i>Treg1 vs. Th1</i>	0.9905	-6.819 to 8.799	ns	0.7957
<i>Treg1 vs. Th2</i>	-6.422	-14.23 to 1.387	ns	0.1026
<i>Treg1 vs. Th1/17</i>	0.4040	-7.405 to 8.213	ns	0.9159
<i>Treg1 vs. Th17</i>	2.304	-5.505 to 10.11	ns	0.5483
<i>Treg1 vs. Th22</i>	2.847	-4.962 to 10.66	ns	0.4590
<i>Treg1 vs. Th9</i>	0.09392	-7.715 to 7.903	ns	0.9804
<i>Treg2 vs. Treg1/17</i>	2.244	-5.565 to 10.05	ns	0.5587
<i>Treg2 vs. Treg17</i>	9.399	1.590 to 17.21	*	0.0204
<i>Treg2 vs. Treg22</i>	5.027	-2.782 to 12.84	ns	0.1965
<i>Treg2 vs. Treg9</i>	10.02	2.211 to 17.83	*	0.0141
<i>Treg2 vs. Th1</i>	10.17	2.360 to 17.98	*	0.0129
<i>Treg2 vs. Th2</i>	2.757	-5.052 to 10.57	ns	0.4733
<i>Treg2 vs. Th1/17</i>	9.582	1.773 to 17.39	*	0.0183
<i>Treg2 vs. Th17</i>	11.48	3.673 to 19.29	**	0.0057
<i>Treg2 vs. Th22</i>	12.03	4.217 to 19.83	**	0.0040
<i>Treg2 vs. Th9</i>	9.272	1.463 to 17.08	*	0.0219
<i>Treg1/17 vs. Treg17</i>	7.155	-0.6536 to 14.96	ns	0.0707
<i>Treg1/17 vs. Treg22</i>	2.783	-5.026 to 10.59	ns	0.4692
<i>Treg1/17 vs. Treg9</i>	7.776	-0.03318 to 15.58	ns	0.0509
<i>Treg1/17 vs. Th1</i>	7.925	0.1160 to 15.73	*	0.0469
<i>Treg1/17 vs. Th2</i>	0.5127	-7.296 to 8.322	ns	0.8933
<i>Treg1/17 vs. Th1/17</i>	7.339	-0.4705 to 15.15	ns	0.0643
<i>Treg1/17 vs. Th17</i>	9.239	1.430 to 17.05	*	0.0224
<i>Treg1/17 vs. Th22</i>	9.782	1.973 to 17.59	*	0.0162
<i>Treg1/17 vs. Th9</i>	7.028	-0.7806 to 14.84	ns	0.0755
<i>Treg17 vs. Treg22</i>	-4.373	-12.18 to 3.436	ns	0.2592
<i>Treg17 vs. Treg9</i>	0.6204	-7.189 to 8.429	ns	0.8711
<i>Treg17 vs. Th1</i>	0.7696	-7.039 to 8.579	ns	0.8405
<i>Treg17 vs. Th2</i>	-6.643	-14.45 to 1.166	ns	0.0919
<i>Treg17 vs. Th1/17</i>	0.1831	-7.626 to 7.992	ns	0.9618
<i>Treg17 vs. Th17</i>	2.083	-5.726 to 9.892	ns	0.5870
<i>Treg17 vs. Th22</i>	2.627	-5.182 to 10.44	ns	0.4942
<i>Treg17 vs. Th9</i>	-0.1270	-7.936 to 7.682	ns	0.9735
<i>Treg22 vs. Treg9</i>	4.993	-2.816 to 12.80	ns	0.1994
<i>Treg22 vs. Th1</i>	5.142	-2.667 to 12.95	ns	0.1868
<i>Treg22 vs. Th2</i>	-2.270	-10.08 to 5.539	ns	0.5541
<i>Treg22 vs. Th1/17</i>	4.556	-3.253 to 12.36	ns	0.2403
<i>Treg22 vs. Th17</i>	6.456	-1.353 to 14.26	ns	0.1009
<i>Treg22 vs. Th22</i>	6.999	-0.8099 to 14.81	ns	0.0767
<i>Treg22 vs. Th9</i>	4.246	-3.563 to 12.05	ns	0.2729
<i>Treg9 vs. Th1</i>	0.1492	-7.660 to 7.958	ns	0.9689
<i>Treg9 vs. Th2</i>	-7.263	-15.07 to 0.5459	ns	0.0669
<i>Treg9 vs. Th1/17</i>	-0.4373	-8.246 to 7.372	ns	0.9090
<i>Treg9 vs. Th17</i>	1.463	-6.346 to 9.272	ns	0.7024
<i>Treg9 vs. Th22</i>	2.006	-5.803 to 9.815	ns	0.6008
<i>Treg9 vs. Th9</i>	-0.7474	-8.556 to 7.062	ns	0.8451
<i>Th1 vs. Th2</i>	-7.412	-15.22 to 0.3967	ns	0.0618
<i>Th1 vs. Th1/17</i>	-0.5865	-8.395 to 7.223	ns	0.8781
<i>Th1 vs. Th17</i>	1.314	-6.495 to 9.123	ns	0.7315
<i>Th1 vs. Th22</i>	1.857	-5.952 to 9.666	ns	0.6280
<i>Th1 vs. Th9</i>	-0.8966	-8.706 to 6.912	ns	0.8147
<i>Th2 vs. Th1/17</i>	6.826	-0.9832 to 14.63	ns	0.0838

<i>Th2 vs. Th17</i>	8.726	0.9169 to 16.53	*	0.0300
<i>Th2 vs. Th22</i>	9.269	1.460 to 17.08	*	0.0220
<i>Th2 vs. Th9</i>	6.516	-1.293 to 14.32	ns	0.0979
<i>Th1/17 vs. Th17</i>	1.900	-5.909 to 9.709	ns	0.6201
<i>Th1/17 vs. Th22</i>	2.443	-5.366 to 10.25	ns	0.5245
<i>Th1/17 vs. Th9</i>	-0.3101	-8.119 to 7.499	ns	0.9354
<i>Th17 vs. Th22</i>	0.5433	-7.266 to 8.352	ns	0.8870
<i>Th17 vs. Th9</i>	-2.210	-10.02 to 5.599	ns	0.5646
<i>Th22 vs. Th9</i>	-2.753	-10.56 to 5.056	ns	0.4738

### 3.4.6 ZEB2 Expression in CM and EM of Purified Tconv Helper Lineage Populations

#### Characterising Central and Effector memory of Tconv helper lineage populations

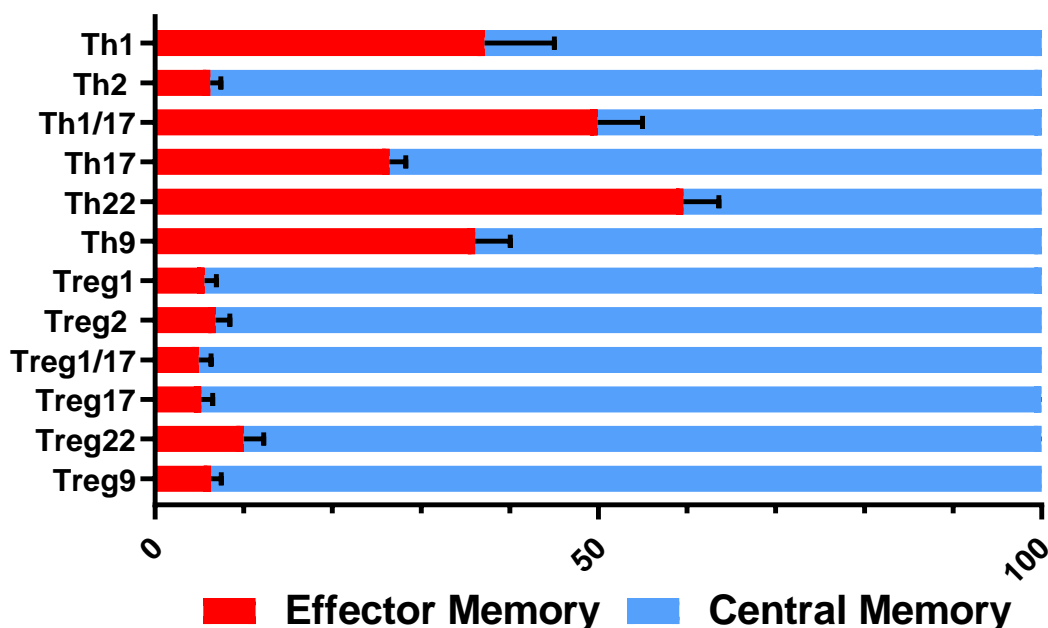
As shown in Figure 3.7 bulk Treg and Tconv are comprised of Central Memory (CM) and Effector Memory (EM) populations, and hence helper lineage populations of Treg and Tconv (Figure 3.12) were able to be further subdivided into CM and EM. The next thing I wished to determine was if there was segregation of ZEB2 expression between CM and EM populations in all of the helper lineage populations, especially in the Th1 where ZEB2 was expressed highly. I separated CD45RA- memory Treg and Tconv with the chemokine receptor panel in combination with the surface marker CD62L, to examine the proportion of CM and EM cells in each of the helper lineage populations. Figure 3.15 shows representative data of the gating strategy for each of the lineages in Treg and Tconv, separating CM and EM by CD62L status and the proportions of CM compared with EM for each of Treg and Tconv helper lineage populations are shown in Figure 3.16. This figure additionally illustrates that Treg have a larger proportion of CM compared with the same lineages within Tconv. Thus, we can see that Treg helper lineage populations are primarily composed of CM cells with only 5-10% EM cells. This explains why the expression ZEB2 in these Treg helper lineage populations (mixed of CM and EM) was low in Figure 3.14, and yet in Figure 3.8, ZEB2 expression in purified Treg EM was higher, compared with purified Tconv EM. Since Tregs are mostly Central Memory, high ZEB2 expression in the Treg Effector Memory is essentially reduced when analyses are carried out on the physiologically representative pooled memory (CM+EM) Treg helper lineage populations. In contrast, the Tconv EM helper lineage populations generally comprise 40%-60% of the total memory cell population, so there is much less impact of Tconv CM in the mixture. There were also differences in the proportion of CM and EM cells in each of the Tconv lineage subsets with the most marked being the Tconv Th2 subset which is almost entirely made up of CM cells (Figure 3.15 and Figure 3.16).



**Figure 3.15: Gating strategy of Central Memory and Effector memory from Treg and Tconv helper lineage populations.**

Histogram showing Tconv (left) and Treg (right) helper lineage populations gated on CD62L expression into Effector Memory (EM) and Central Memory (CM). Enriched memory CD4 were labelled with surface antibodies from panel 4 (Table 3.4) including CD4, CD25, CD127, CD45RA, CXCR3, CCR6, CCR4, CCR10 and CD62L. Cells were first physically gated on singlets, lymphocytes, memory CD4<sup>+</sup> and further divided into memory Treg and memory Tconv before separating into different helper lineages based on chemokine receptors expression. A gate was then set to divide CD62L expression to CD62L<sup>-</sup> (EM) and CD62L<sup>+</sup> (CM) cells from each helper lineage populations of Treg and Tconv. Representative data from 1 donor, n=3 independent donors.





**Figure 3.16: Proportion of Effector Memory and Central Memory subsets in Tconv and Treg helper lineage populations**

Bar graph showing the proportion of Effector Memory (red) and Central Memory (blue) subsets of Tconv and Treg helper lineage populations generated with mean and SEM indicated in Table 3.18 from 5 donors.

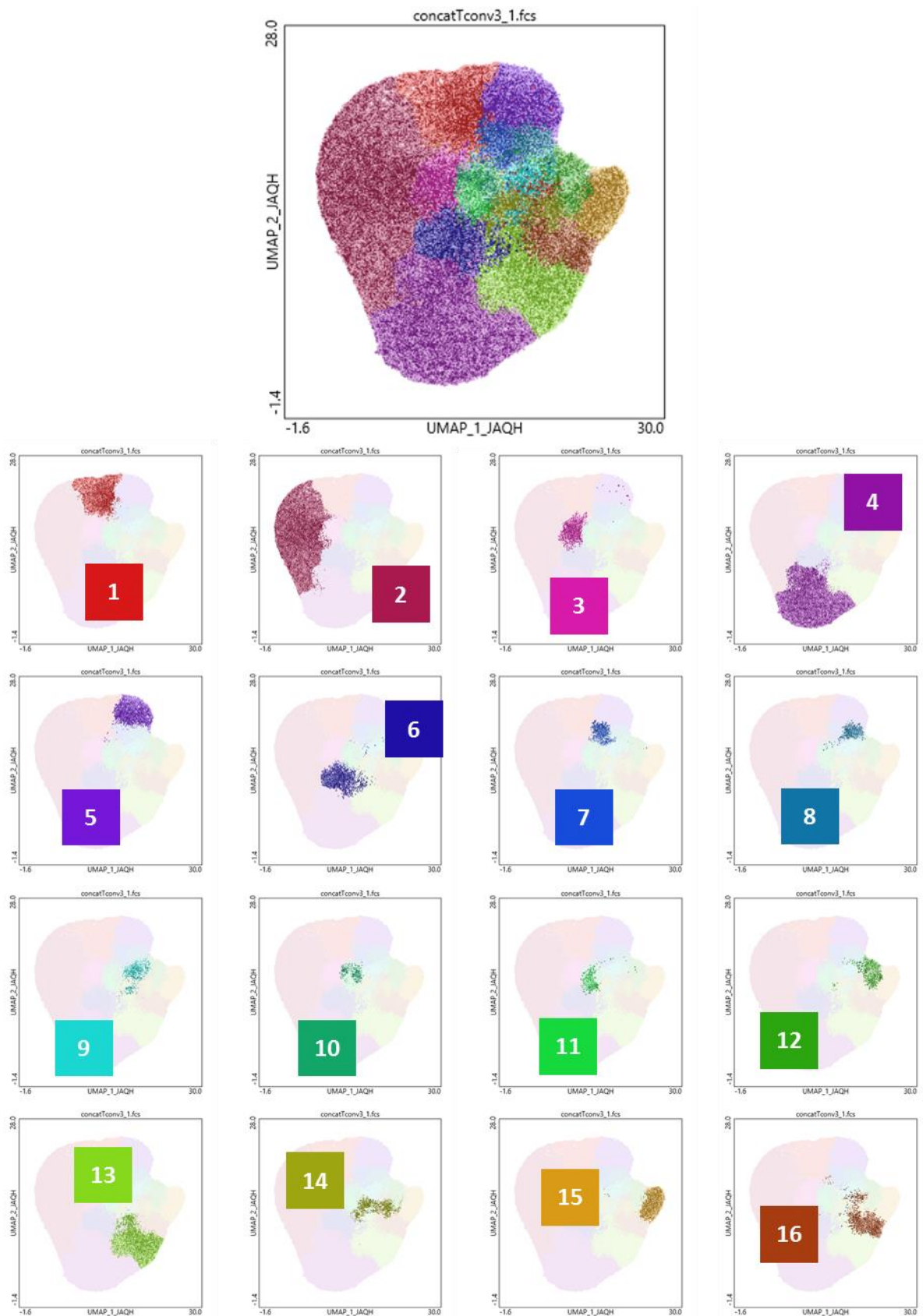
**Table 3.18: Mean and SEM of Effector Memory and Central Memory subsets proportion in each Tconv and Treg helper lineage populations**

Populations	Effector Memory			Central Memory		
	Mean	SEM	N	Mean	SEM	N
Th1	37.2	7.859071	5	62.8	7.859071	5
Th2	6.26	1.184707	5	93.74	1.185158	5
Th1/17	49.92	5.056026	5	50.08	5.056026	5
Th17	26.442	1.82876	5	73.56	1.829371	5
Th22	59.62	4.004672	5	40.42	3.992418	5
Th9	36.1	3.980327	5	63.9	3.980327	5
Treg1	5.6	1.318169	5	94.4	1.311488	5
Treg2	6.834	1.62472	5	93.18	1.624007	5
Treg1/17	4.956	1.376033	5	95.06	1.377897	5
Treg17	5.248	1.251405	5	94.72	1.25036	5
Treg22	10.014	2.276681	5	89.98	2.272972	5
Treg9	6.346	1.148619	5	93.66	1.149174	5

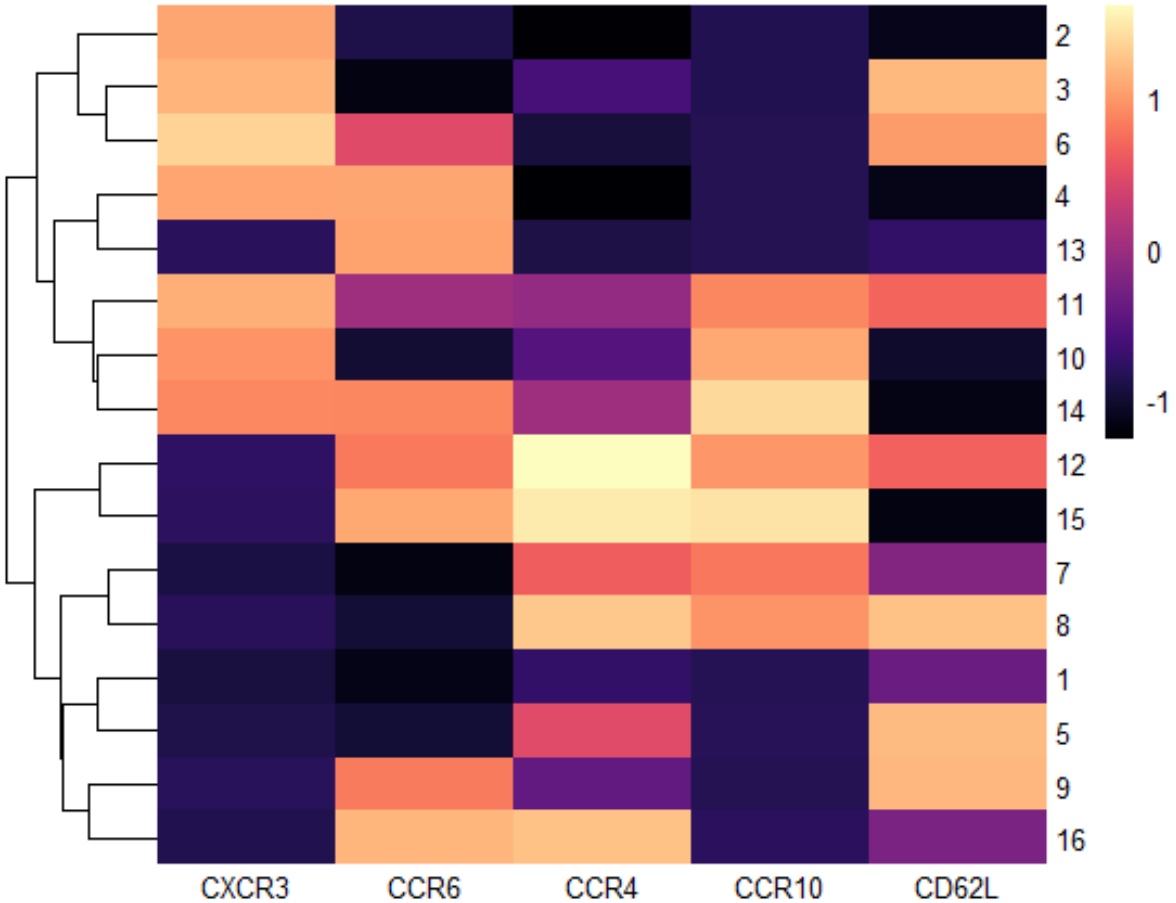
### Dimension reduction of Memory Tconv pools with UMAP

In order to validate the veracity of using the chemokine receptor panel and memory subset panel to correctly identify purified populations in an unbiased way, Uniform Manifold Approximation and Projection (UMAP) analysis was performed on the FCS files acquired during the sort to group cells with similar surface marker expression and projected in two dimensions. Memory Tconv cells that were stained with the chemokine receptor panel and the memory subset panel from three individual donors were merged to cluster the cells based on similar surface marker expression. A total of 16 clusters were identified from the UMAP (Figure 3.17), and the classification of the clusters shown in Table 3.19 were based on chemokine receptor expression profiles as indicated in Table 3.4. Here, clusters 2 and 3 were of particular interest as these appear to be Th1 owing to high expression of CXCR3 alone, where CD62L expression dictates the EM (cluster 2) and CM (cluster 3) status (Figure 3.18 & Table 3.19). Cluster 2 Th1 EM is the largest cluster compared with the others. Interestingly the UMAP also captures some populations that have not been identified using our sorting strategy. These include clusters 1, 7, 8, 10, 11 and 14 (Table 3.19). The ThNeg population from cluster 1 for example, which has negative expression for all four chemokine receptors and CD62L surface markers, is 8% of the memory Tconv population. It is quite likely that additional chemokine receptors such as CXCR5 may be used to tease out T follicular helper (Tfh) cells from this cluster to examine ZEB2 expression in the future (Table 3.19). Based on their chemokine receptor expression, Clusters 7 and 8 are possibly Th2/22 but require more validation and therefore not included in my analysis. Clusters 10, 11 and 14 may be novel populations, although their chemokine receptor expression patterns suggest that they may be subsets of Th22 that have been described by Eyerich et al. (2009) and Kuang et al. (2014). In this paper, Kuang et al. (2014) describe 8 non-classical Th22 subsets that secrete combinations of IFN $\gamma$ , IL-4 and/or IL-17 thus sharing some characteristics of Th1, Th2 and Th17 cells.

However, further analyses of these Th22 subsets was beyond the scope of my PhD project. Some well characterised populations of cells were not detected using this analytical tool including Th17 CM and Th2 EM. This may be owing to expression patterns very similar to other clusters, or for the requirement of higher cell numbers to build a higher resolution UMAP that could capture rare cell populations. Overall, the gating strategy I used captured more than 60% of the populations detected in UMAP, suggesting high confidence of our purified population classification.



**Figure 3.17: UMAP of memory Tconv stained with the chemokine receptor panel**  
 UMAP embedding of merged profiles of 3 donors with each containing 41260 cells. UMAP embeddings are coloured by expression cluster from 1 to 16.



**Figure 3.18: MFI Heatmap of UMAP clusters**  
Heatmap showing the relative intensity of chemokine receptors expression normalised by z-score for UMAP clusters from merged profiles of 3 donors.

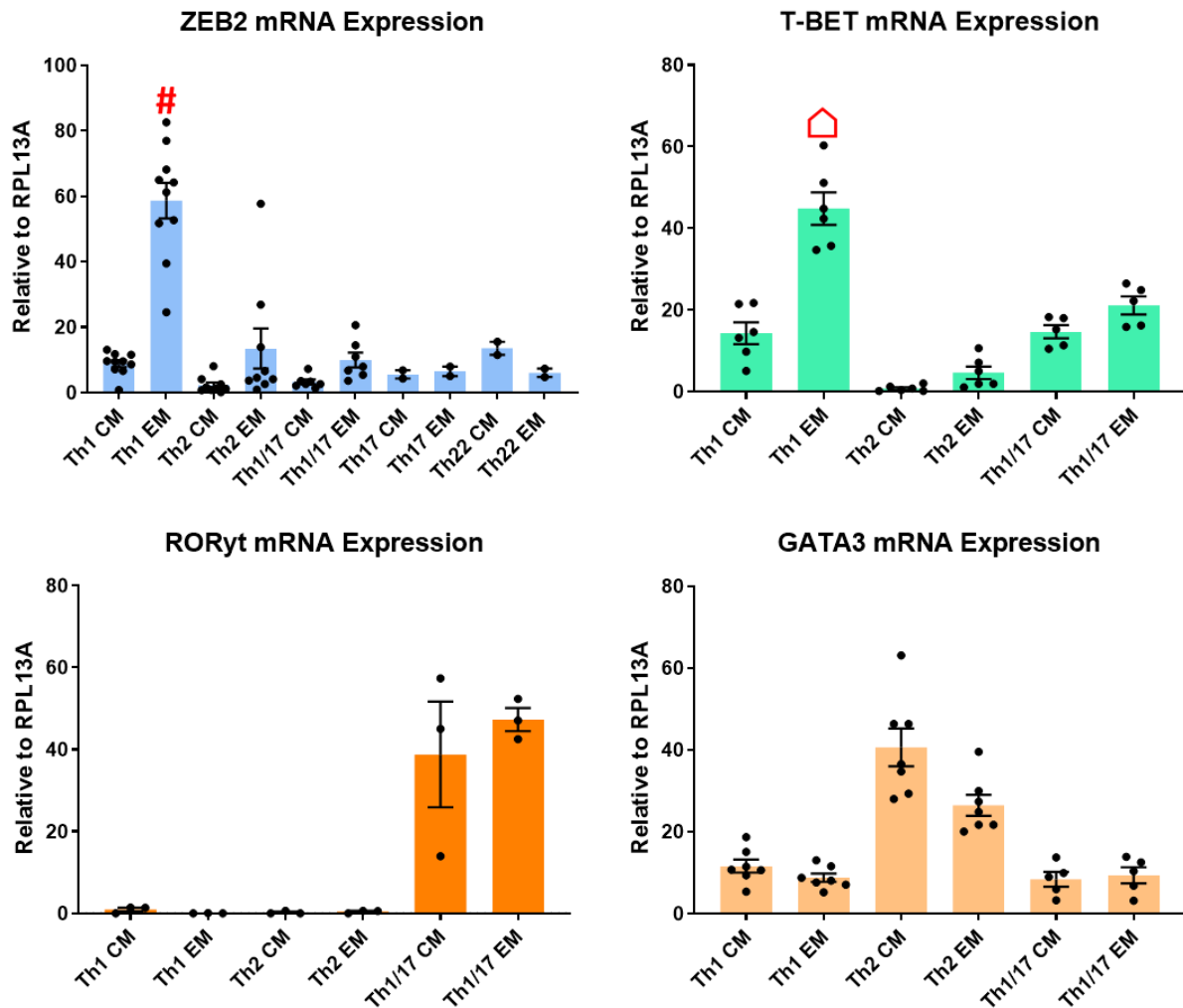
**Table 3.19: Proportion of UMAP clusters with population identity classified by chemokine receptor expression**

Cluster	Population	Proportion (%)
1	ThNeg EM	8.78
2	Th1 EM	29.1
3	Th1 CM	2.19
4	Th1/17 EM	18
5	Th2 CM	6.67
6	Th1/17 CM	4
7	Th2/22 EM	1.17
8	Th2/22 EM	0.62
9	Th9 CM	5.07
10	CXCR3+ CCR10+ EM	1.08
11	CXCR3+ CCR10+ CM	0.39
12	Th22 CM	1.03
13	Th9 EM	12.4
14	CXCR3+ CCR6+ CCR10+ EM	1.13
15	Th22 EM	5.07
16	Th17 EM	3.32

#### ZEB2 expression across memory subsets of Tconv helper lineage populations

To determine if ZEB2 is expressed in both CM and EM of Th1 or if ZEB2 is lineage and maturation restricted, total RNAs were isolated from CM and EM of Tconv helper lineages (depicted in Figure 3.15). I omitted the Treg CM and EM helper lineages since they were extremely low abundance and I had already shown that ZEB2 mRNA expression was very low in these cells (Figure 3.14). ZEB2 expression, together with the Th1 lineage defining transcription factor T-bet, the Th1/17 lineage defining transcription factor ROR $\gamma$ t and the Th2 lineage defining transcription factor GATA3, were examined for comparison (Figure 3.19). T-bet was expressed predominantly in the Th1 EM population, with significantly lower expression

in the Central Memory population and this was echoed by the expression of ZEB2. As mentioned above, T-bet is also found in the Th1/17 cells, and with this further interrogation I showed that T-bet is expressed in both the CM and EM of Th1/17 cells. The Th17 transcription factor ROR $\gamma$ t, is found in both CM and EM Th1/17 and the Th2 defining transcription factor, GATA3, is expressed in both the Th2 Central and EM populations. Measuring expression of these lineage defining transcription factors also acts as a quality control measure to ensure that our gating strategy is well defined in correctly identifying Tconv helper populations. These experiments were carried out with great rigour: cells were isolated and analysed from 5 different donors using very strict, careful, and consistent gating strategies and the data was carefully validated using multiple strategies. Therefore, I am confident to assert that this data shows that ZEB2 mRNA expression is almost exclusively confined to Th1 EM cells. Thus, my combined comprehensive phenotyping and expression analysis reveals that the role of ZEB2 is highly likely to be restricted to the Th1 EM cell population. This could be potentially for promoting or reinforcing the Th1 effector function and /or for the formation of Th1 or maintaining Th1 lineage fidelity.



**Figure 3.19: ZEB2, T-bet, GATA3 and RORγt mRNA expression in memory subsets of Tconv helper lineage population**

ZEB2 (blue), T-bet (green), GATA3 (orange) and RORγt (red) expression in memory subsets of Tconv helper lineage populations. Relative abundance of ZEB2, T-bet, GATA3 and, RORγt were normalised to reference gene RPL13A and plotted with mean + SEM. As there are multiple comparisons, the statistical significance is presented in a separate table, to keep the figures clear. The annotation # indicates the level of ZEB2 in Th1 EM is significantly higher than other populations whereas the annotation Δ indicates the level of T-bet is significantly higher than other populations. Statistics were carried out with ordinary one-way Anova multiple comparison Fisher's Least Significant Difference (LSD) test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$  as shown in Table 3.20, Table 3.21, Table 3.22 and Table 3.23,  $n = 3-10$  independent donors.



**Table 3.20: Anova Fisher's LSD test on ZEB2 expression in memory subsets of Tconv helper lineage populations**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM vs. Th1 EM</i>	-49.82	-59.50 to -40.13	****	<0.0001
<i>Th1 CM vs. Th2 CM</i>	6.558	-3.396 to 16.51	ns	0.1917
<i>Th1 CM vs. Th2 EM</i>	-4.611	-14.56 to 5.343	ns	0.3566
<i>Th1 CM vs. Th1/17 CM</i>	5.548	-5.127 to 16.22	ns	0.3016
<i>Th1 CM vs. Th1/17 EM</i>	-1.126	-11.80 to 9.550	ns	0.8331
<i>Th1 CM vs. Th17 CM</i>	3.269	-13.51 to 20.05	ns	0.6972
<i>Th1 CM vs. Th17 EM</i>	2.337	-14.44 to 19.12	ns	0.7809
<i>Th1 CM vs. Th22 CM</i>	-4.686	-21.47 to 12.09	ns	0.5774
<i>Th1 CM vs. Th22 EM</i>	2.795	-13.99 to 19.58	ns	0.7393
<i>Th1 EM vs. Th2 CM</i>	56.37	46.42 to 66.33	****	<0.0001
<i>Th1 EM vs. Th2 EM</i>	45.20	35.25 to 55.16	****	<0.0001
<i>Th1 EM vs. Th1/17 CM</i>	55.36	44.69 to 66.04	****	<0.0001
<i>Th1 EM vs. Th1/17 EM</i>	48.69	38.01 to 59.37	****	<0.0001
<i>Th1 EM vs. Th17 CM</i>	53.09	36.31 to 69.87	****	<0.0001
<i>Th1 EM vs. Th17 EM</i>	52.15	35.37 to 68.93	****	<0.0001
<i>Th1 EM vs. Th22 CM</i>	45.13	28.35 to 61.91	****	<0.0001
<i>Th1 EM vs. Th22 EM</i>	52.61	35.83 to 69.39	****	<0.0001
<i>Th2 CM vs. Th2 EM</i>	-11.17	-21.38 to -0.9569	*	0.0327
<i>Th2 CM vs. Th1/17 CM</i>	-1.010	-11.93 to 9.908	ns	0.8534
<i>Th2 CM vs. Th1/17 EM</i>	-7.684	-18.60 to 3.234	ns	0.1637
<i>Th2 CM vs. Th17 CM</i>	-3.289	-20.22 to 13.65	ns	0.6982
<i>Th2 CM vs. Th17 EM</i>	-4.221	-21.16 to 12.71	ns	0.6188
<i>Th2 CM vs. Th22 CM</i>	-11.24	-28.18 to 5.691	ns	0.1884
<i>Th2 CM vs. Th22 EM</i>	-3.763	-20.70 to 13.17	ns	0.6573
<i>Th2 EM vs. Th1/17 CM</i>	10.16	-0.7578 to 21.08	ns	0.0675
<i>Th2 EM vs. Th1/17 EM</i>	3.485	-7.432 to 14.40	ns	0.5243
<i>Th2 EM vs. Th17 CM</i>	7.880	-9.054 to 24.82	ns	0.3545
<i>Th2 EM vs. Th17 EM</i>	6.948	-9.987 to 23.88	ns	0.4138
<i>Th2 EM vs. Th22 CM</i>	-0.07484	-17.01 to 16.86	ns	0.9930
<i>Th2 EM vs. Th22 EM</i>	7.406	-9.529 to 24.34	ns	0.3839
<i>Th1/17 CM vs. Th1/17 EM</i>	-6.674	-18.25 to 4.905	ns	0.2525
<i>Th1/17 CM vs. Th17 CM</i>	-2.279	-19.65 to 15.09	ns	0.7932
<i>Th1/17 CM vs. Th17 EM</i>	-3.212	-20.58 to 14.16	ns	0.7119
<i>Th1/17 CM vs. Th22 CM</i>	-10.23	-27.60 to 7.135	ns	0.2422
<i>Th1/17 CM vs. Th22 EM</i>	-2.753	-20.12 to 14.62	ns	0.7515
<i>Th1/17 EM vs. Th17 CM</i>	4.395	-12.97 to 21.76	ns	0.6135
<i>Th1/17 EM vs. Th17 EM</i>	3.462	-13.91 to 20.83	ns	0.6906
<i>Th1/17 EM vs. Th22 CM</i>	-3.560	-20.93 to 13.81	ns	0.6823
<i>Th1/17 EM vs. Th22 EM</i>	3.921	-13.45 to 21.29	ns	0.6522
<i>Th17 CM vs. Th17 EM</i>	-0.9327	-22.60 to 20.73	ns	0.9314
<i>Th17 CM vs. Th22 CM</i>	-7.955	-29.62 to 13.71	ns	0.4642
<i>Th17 CM vs. Th22 EM</i>	-0.4742	-22.14 to 21.19	ns	0.9651
<i>Th17 EM vs. Th22 CM</i>	-7.023	-28.69 to 14.64	ns	0.5180
<i>Th17 EM vs. Th22 EM</i>	0.4585	-21.20 to 22.12	ns	0.9663
<i>Th22 CM vs. Th22 EM</i>	7.481	-14.18 to 29.14	ns	0.4911

**Table 3.21: Anova Fisher's LSD test on T-bet expression in memory subsets of Tconv helper lineage populations**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM vs. Th1 EM</i>	-30.54	-37.26 to -23.81	****	<0.0001
<i>Th1 CM vs. Th2 CM</i>	13.52	6.794 to 20.25	***	0.0003
<i>Th1 CM vs. Th2 EM</i>	9.693	2.965 to 16.42	**	0.0063
<i>Th1 CM vs. Th1/17 CM</i>	-0.3672	-7.423 to 6.689	ns	0.9159
<i>Th1 CM vs. Th1/17 EM</i>	-6.818	-13.87 to 0.2381	ns	0.0577
<i>Th1 EM vs. Th2 CM</i>	44.06	37.33 to 50.79	****	<0.0001
<i>Th1 EM vs. Th2 EM</i>	40.23	33.50 to 46.96	****	<0.0001
<i>Th1 EM vs. Th1/17 CM</i>	30.17	23.11 to 37.22	****	<0.0001
<i>Th1 EM vs. Th1/17 EM</i>	23.72	16.66 to 30.77	****	<0.0001
<i>Th2 CM vs. Th2 EM</i>	-3.829	-10.56 to 2.899	ns	0.2535
<i>Th2 CM vs. Th1/17 CM</i>	-13.89	-20.94 to -6.833	***	0.0004
<i>Th2 CM vs. Th1/17 EM</i>	-20.34	-27.40 to -13.28	****	<0.0001
<i>Th2 EM vs. Th1/17 CM</i>	-10.06	-17.12 to -3.004	**	0.0068
<i>Th2 EM vs. Th1/17 EM</i>	-16.51	-23.57 to -9.455	****	<0.0001
<i>Th1/17 CM vs. Th1/17 EM</i>	-6.450	-13.82 to 0.9191	ns	0.0838

**Table 3.22: Anova Fisher's LSD test on ROR $\gamma$ t expression in memory subsets of Tconv helper lineage populations**

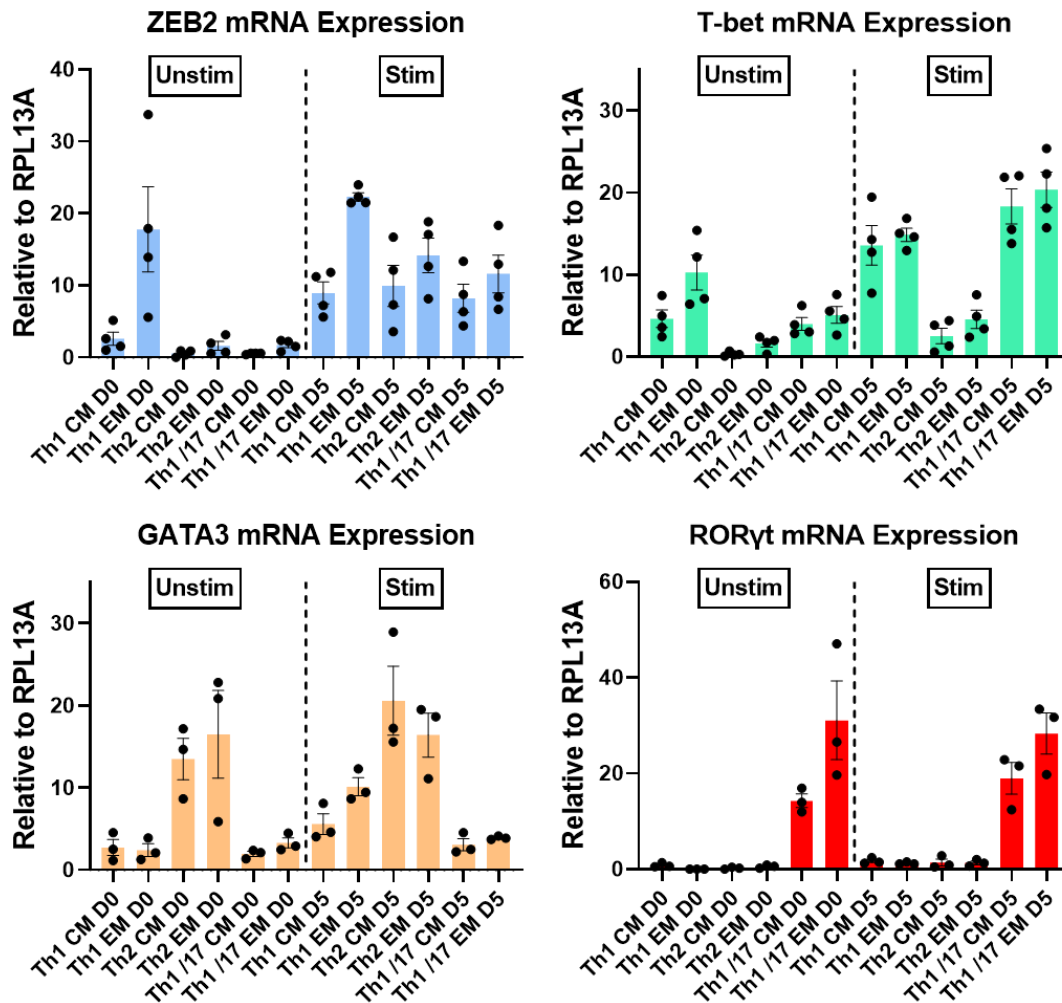
Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM vs. Th1 EM</i>	0.9130	-15.73 to 17.56	ns	0.9069
<i>Th1 CM vs. Th2 CM</i>	0.7171	-15.93 to 17.36	ns	0.9268
<i>Th1 CM vs. Th2 EM</i>	0.5271	-16.12 to 17.17	ns	0.9461
<i>Th1 CM vs. Th1/17 CM</i>	-37.85	-54.50 to -21.21	***	0.0003
<i>Th1 CM vs. Th1/17 EM</i>	-46.39	-63.04 to -29.75	****	<0.0001
<i>Th1 EM vs. Th2 CM</i>	-0.1959	-16.84 to 16.45	ns	0.9800
<i>Th1 EM vs. Th2 EM</i>	-0.3859	-17.03 to 16.26	ns	0.9605
<i>Th1 EM vs. Th1/17 CM</i>	-38.77	-55.41 to -22.12	***	0.0003
<i>Th1 EM vs. Th1/17 EM</i>	-47.31	-63.95 to -30.66	****	<0.0001
<i>Th2 CM vs. Th2 EM</i>	-0.1900	-16.84 to 16.46	ns	0.9806
<i>Th2 CM vs. Th1/17 CM</i>	-38.57	-55.22 to -21.92	***	0.0003
<i>Th2 CM vs. Th1/17 EM</i>	-47.11	-63.76 to -30.46	****	<0.0001
<i>Th2 EM vs. Th1/17 CM</i>	-38.38	-55.03 to -21.73	***	0.0003
<i>Th2 EM vs. Th1/17 EM</i>	-46.92	-63.57 to -30.27	****	<0.0001
<i>Th1/17 CM vs. Th1/17 EM</i>	-8.540	-25.19 to 8.108	ns	0.2856

**Table 3.23: Anova Fisher's LSD test on GATA3 expression in memory subsets of Tconv helper lineage populations**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM vs. Th1 EM</i>	2.856	-4.535 to 10.25	ns	0.4370
<i>Th1 CM vs. Th2 CM</i>	-29.01	-36.40 to -21.62	****	<0.0001
<i>Th1 CM vs. Th2 EM</i>	-14.83	-22.22 to -7.443	***	0.0003
<i>Th1 CM vs. Th1/17 CM</i>	3.249	-4.847 to 11.35	ns	0.4197
<i>Th1 CM vs. Th1/17 EM</i>	2.263	-5.833 to 10.36	ns	0.5731
<i>Th1 EM vs. Th2 CM</i>	-31.87	-39.26 to -24.48	****	<0.0001
<i>Th1 EM vs. Th2 EM</i>	-17.69	-25.08 to -10.30	****	<0.0001
<i>Th1 EM vs. Th1/17 CM</i>	0.3934	-7.702 to 8.489	ns	0.9218
<i>Th1 EM vs. Th1/17 EM</i>	-0.5928	-8.689 to 7.503	ns	0.8824
<i>Th2 CM vs. Th2 EM</i>	14.18	6.785 to 21.57	***	0.0005
<i>Th2 CM vs. Th1/17 CM</i>	32.26	24.16 to 40.35	****	<0.0001
<i>Th2 CM vs. Th1/17 EM</i>	31.27	23.18 to 39.37	****	<0.0001
<i>Th2 EM vs. Th1/17 CM</i>	18.08	9.987 to 26.18	****	<0.0001
<i>Th2 EM vs. Th1/17 EM</i>	17.10	9.001 to 25.19	***	0.0001
<i>Th1/17 CM vs. Th1/17 EM</i>	-0.9862	-9.731 to 7.758	ns	0.8198

### 3.4.7 Characterising ZEB2 expression in CM and EM following activation

I have shown that ZEB2 is almost exclusively expressed in Th1 EM, but in order to understand the role of ZEB2 in the Th1 EM, I wanted to characterise its expression further. For this reason, I decided to examine ZEB2 expression in Th1 cells after activation. Central and Effector memory populations were isolated and activated (1:1 anti-CD3/CD28 magnetic bead, Dynabeads® Human T-Activator CD3/CD28, Life Technologies) for 5 days (as indicated in Section 3.3.6) and RNA was then collected. As can be seen in Figure 3.20, in non-activated cells ZEB2 is only found in the Th1 EM cell lineage and this high expression of ZEB2 in the Th1 EM was maintained in populations of cells with or without activation, suggesting that a high basal level of ZEB2 is required by these cells at all times. However, although very little ZEB2 was found in non-activated Th2, Th1/17 CM and EM cells, after 5 days of activation ZEB2 expression in each of the other lineages increases. This could mean that there is an activation component to ZEB2 induction in EM cells. Interestingly, this is likewise observed in the CM lineages. There is very little ZEB2 in any of the CM subsets in non-activated cells, including Th1 CM cells. However, ZEB2 is induced upon activation in all of these subsets, although not to the same extent as in the EM subsets.



**Figure 3.20: ZEB2 (blue), T-bet (green), GATA3 (orange) and RORyt (red) mRNA expression in non-activated (unstim) or activated (stim) Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th1/17**

Expression of ZEB2 (blue), T-bet (green), GATA3 (orange) and RORyt (red) mRNAs were examined in CM and EM of Th1, Th2, and Th1/17 cells activated with 1:1 anti-CD3/CD28 magnetic beads for 5 days (D5) and compared with cells from unstimulated conditions where cells were freshly isolated on day 0 (D0). Relative abundance of ZEB2, T-bet, GATA3 and RORyt were normalised to reference gene RPL13A and plotted with mean + SEM. As there are multiple comparisons, the statistical significance is presented in a separate table, to keep the figures clear. Statistics were carried out with ordinary one-way Anova multiple comparison Fisher's Least Significant Difference (LSD) test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$  as shown in Table 3.24, Table 3.25, Table 3.26 and Table 3.27,  $n = 3$  independent donors.

**Table 3.24: Anova Fisher's LSD test on ZEB2 expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM D0 vs. Th1 EM D0</i>	-15.21	-21.80 to -8.613	****	<0.0001
<i>Th1 CM D0 vs. Th2 CM D0</i>	2.046	-4.548 to 8.640	ns	0.5332
<i>Th1 CM D0 vs. Th2 EM D0</i>	0.9854	-5.609 to 7.580	ns	0.7636
<i>Th1 CM D0 vs. Th1 /17 CM D0</i>	2.053	-4.542 to 8.647	ns	0.5318
<i>Th1 CM D0 vs. Th1 /17 EM D0</i>	0.8665	-5.728 to 7.461	ns	0.7914
<i>Th1 CM D0 vs. Th1 CM D5</i>	-6.376	-12.97 to 0.2187	ns	0.0577
<i>Th1 CM D0 vs. Th1 EM D5</i>	-19.72	-26.32 to -13.13	****	<0.0001
<i>Th1 CM D0 vs. Th2 CM D5</i>	-7.352	-13.95 to -0.7574	*	0.0299
<i>Th1 CM D0 vs. Th2 EM D5</i>	-11.60	-18.19 to -5.001	**	0.0010
<i>Th1 CM D0 vs. Th1 /17 CM D5</i>	-5.622	-12.22 to 0.9725	ns	0.0924
<i>Th1 CM D0 vs. Th1 /17 EM D5</i>	-9.016	-15.61 to -2.422	**	0.0087
<i>Th1 EM D0 vs. Th2 CM D0</i>	17.25	10.66 to 23.85	****	<0.0001
<i>Th1 EM D0 vs. Th2 EM D0</i>	16.19	9.598 to 22.79	****	<0.0001
<i>Th1 EM D0 vs. Th1 /17 CM D0</i>	17.26	10.67 to 23.85	****	<0.0001
<i>Th1 EM D0 vs. Th1 /17 EM D0</i>	16.07	9.479 to 22.67	****	<0.0001
<i>Th1 EM D0 vs. Th1 CM D5</i>	8.832	2.237 to 15.43	*	0.0101
<i>Th1 EM D0 vs. Th1 EM D5</i>	-4.517	-11.11 to 2.077	ns	0.1733
<i>Th1 EM D0 vs. Th2 CM D5</i>	7.856	1.261 to 14.45	*	0.0209
<i>Th1 EM D0 vs. Th2 EM D5</i>	3.612	-2.982 to 10.21	ns	0.2740
<i>Th1 EM D0 vs. Th1 /17 CM D5</i>	9.585	2.991 to 16.18	**	0.0056
<i>Th1 EM D0 vs. Th1 /17 EM D5</i>	6.191	-0.4033 to 12.79	ns	0.0649
<i>Th2 CM D0 vs. Th2 EM D0</i>	-1.061	-7.655 to 5.534	ns	0.7462
<i>Th2 CM D0 vs. Th1 /17 CM D0</i>	0.006797	-6.587 to 6.601	ns	0.9983
<i>Th2 CM D0 vs. Th1 /17 EM D0</i>	-1.179	-7.774 to 5.415	ns	0.7189
<i>Th2 CM D0 vs. Th1 CM D5</i>	-8.421	-15.02 to -1.827	*	0.0138
<i>Th2 CM D0 vs. Th1 EM D5</i>	-21.77	-28.36 to -15.18	****	<0.0001
<i>Th2 CM D0 vs. Th2 CM D5</i>	-9.397	-15.99 to -2.803	**	0.0065
<i>Th2 CM D0 vs. Th2 EM D5</i>	-13.64	-20.24 to -7.047	***	0.0002
<i>Th2 CM D0 vs. Th1 /17 CM D5</i>	-7.668	-14.26 to -1.073	*	0.0239
<i>Th2 CM D0 vs. Th1 /17 EM D5</i>	-11.06	-17.66 to -4.468	**	0.0017
<i>Th2 EM D0 vs. Th1 /17 CM D0</i>	1.067	-5.527 to 7.662	ns	0.7446
<i>Th2 EM D0 vs. Th1 /17 EM D0</i>	-0.1188	-6.713 to 6.475	ns	0.9710
<i>Th2 EM D0 vs. Th1 CM D5</i>	-7.361	-13.96 to -0.7667	*	0.0297
<i>Th2 EM D0 vs. Th1 EM D5</i>	-20.71	-27.30 to -14.12	****	<0.0001
<i>Th2 EM D0 vs. Th2 CM D5</i>	-8.337	-14.93 to -1.743	*	0.0147
<i>Th2 EM D0 vs. Th2 EM D5</i>	-12.58	-19.17 to -5.986	***	0.0004
<i>Th2 EM D0 vs. Th1 /17 CM D5</i>	-6.607	-13.20 to -0.01290	*	0.0496
<i>Th2 EM D0 vs. Th1 /17 EM D5</i>	-10.00	-16.60 to -3.407	**	0.0040
<i>Th1 /17 CM D0 vs. Th1 /17 EM D0</i>	-1.186	-7.780 to 5.408	ns	0.7174
<i>Th1 /17 CM D0 vs. Th1 CM D5</i>	-8.428	-15.02 to -1.834	*	0.0137
<i>Th1 /17 CM D0 vs. Th1 EM D5</i>	-21.78	-28.37 to -15.18	****	<0.0001
<i>Th1 /17 CM D0 vs. Th2 CM D5</i>	-9.404	-16.00 to -2.810	**	0.0064
<i>Th1 /17 CM D0 vs. Th2 EM D5</i>	-13.65	-20.24 to -7.053	***	0.0002
<i>Th1 /17 CM D0 vs. Th1 /17 CM D5</i>	-7.674	-14.27 to -1.080	*	0.0238
<i>Th1 /17 CM D0 vs. Th1 /17 EM D5</i>	-11.07	-17.66 to -4.475	**	0.0016
<i>Th1 /17 EM D0 vs. Th1 CM D5</i>	-7.242	-13.84 to -0.6478	*	0.0323
<i>Th1 /17 EM D0 vs. Th1 EM D5</i>	-20.59	-27.19 to -14.00	****	<0.0001
<i>Th1 /17 EM D0 vs. Th2 CM D5</i>	-8.218	-14.81 to -1.624	*	0.0160
<i>Th1 /17 EM D0 vs. Th2 EM D5</i>	-12.46	-19.06 to -5.867	***	0.0005
<i>Th1 /17 EM D0 vs. Th1 /17 CM D5</i>	-6.488	-13.08 to 0.1059	ns	0.0536
<i>Th1 /17 EM D0 vs. Th1 /17 EM D5</i>	-9.883	-16.48 to -3.289	**	0.0044
<i>Th1 CM D5 vs. Th1 EM D5</i>	-13.35	-19.94 to -6.755	***	0.0002
<i>Th1 CM D5 vs. Th2 CM D5</i>	-0.9761	-7.570 to 5.618	ns	0.7657
<i>Th1 CM D5 vs. Th2 EM D5</i>	-5.219	-11.81 to 1.375	ns	0.1172
<i>Th1 CM D5 vs. Th1 /17 CM D5</i>	0.7538	-5.840 to 7.348	ns	0.8180
<i>Th1 CM D5 vs. Th1 /17 EM D5</i>	-2.641	-9.235 to 3.954	ns	0.4220

<i>Th1 EM D5 vs. Th2 CM D5</i>	12.37	5.778 to 18.97	***	0.0005
<i>Th1 EM D5 vs. Th2 EM D5</i>	8.129	1.535 to 14.72	*	0.0171
<i>Th1 EM D5 vs. Th1 /17 CM D5</i>	14.10	7.508 to 20.70	***	0.0001
<i>Th1 EM D5 vs. Th1 /17 EM D5</i>	10.71	4.114 to 17.30	**	0.0022
<i>Th2 CM D5 vs. Th2 EM D5</i>	-4.243	-10.84 to 2.351	ns	0.2001
<i>Th2 CM D5 vs. Th1 /17 CM D5</i>	1.730	-4.864 to 8.324	ns	0.5980
<i>Th2 CM D5 vs. Th1 /17 EM D5</i>	-1.665	-8.259 to 4.930	ns	0.6118
<i>Th2 EM D5 vs. Th1 /17 CM D5</i>	5.973	-0.6210 to 12.57	ns	0.0745
<i>Th2 EM D5 vs. Th1 /17 EM D5</i>	2.579	-4.015 to 9.173	ns	0.4329
<i>Th1 /17 CM D5 vs. Th1 /17 EM D5</i>	-3.394	-9.989 to 3.200	ns	0.3035



**Table 3.25: Anova Fisher's LSD test on T-bet expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM D0 vs. Th1 EM D0</i>	-5.640	-9.814 to -1.466	**	0.0095
<i>Th1 CM D0 vs. Th2 CM D0</i>	4.300	0.1258 to 8.474	*	0.0438
<i>Th1 CM D0 vs. Th2 EM D0</i>	3.005	-1.169 to 7.179	ns	0.1530
<i>Th1 CM D0 vs. Th1 /17 CM D0</i>	0.6287	-3.545 to 4.803	ns	0.7618
<i>Th1 CM D0 vs. Th1 /17 EM D0</i>	-0.4776	-4.652 to 3.696	ns	0.8178
<i>Th1 CM D0 vs. Th1 CM D5</i>	-8.940	-13.11 to -4.766	***	0.0001
<i>Th1 CM D0 vs. Th1 EM D5</i>	-10.24	-14.41 to -6.063	****	<0.0001
<i>Th1 CM D0 vs. Th2 CM D5</i>	2.093	-2.081 to 6.267	ns	0.3160
<i>Th1 CM D0 vs. Th2 EM D5</i>	0.05196	-4.122 to 4.226	ns	0.9800
<i>Th1 CM D0 vs. Th1 /17 CM D5</i>	-13.68	-17.86 to -9.509	****	<0.0001
<i>Th1 CM D0 vs. Th1 /17 EM D5</i>	-15.74	-19.91 to -11.57	****	<0.0001
<i>Th1 EM D0 vs. Th2 CM D0</i>	9.940	5.766 to 14.11	****	<0.0001
<i>Th1 EM D0 vs. Th2 EM D0</i>	8.645	4.471 to 12.82	***	0.0002
<i>Th1 EM D0 vs. Th1 /17 CM D0</i>	6.269	2.095 to 10.44	**	0.0043
<i>Th1 EM D0 vs. Th1 /17 EM D0</i>	5.163	0.9886 to 9.337	*	0.0168
<i>Th1 EM D0 vs. Th1 CM D5</i>	-3.300	-7.474 to 0.8742	ns	0.1176
<i>Th1 EM D0 vs. Th1 EM D5</i>	-4.597	-8.771 to -0.4230	*	0.0318
<i>Th1 EM D0 vs. Th2 CM D5</i>	7.733	3.559 to 11.91	***	0.0006
<i>Th1 EM D0 vs. Th2 EM D5</i>	5.692	1.518 to 9.866	**	0.0089
<i>Th1 EM D0 vs. Th1 /17 CM D5</i>	-8.043	-12.22 to -3.869	***	0.0004
<i>Th1 EM D0 vs. Th1 /17 EM D5</i>	-10.10	-14.27 to -5.925	****	<0.0001
<i>Th2 CM D0 vs. Th2 EM D0</i>	-1.295	-5.469 to 2.879	ns	0.5331
<i>Th2 CM D0 vs. Th1 /17 CM D0</i>	-3.671	-7.845 to 0.5029	ns	0.0829
<i>Th2 CM D0 vs. Th1 /17 EM D0</i>	-4.777	-8.951 to -0.6034	*	0.0260
<i>Th2 CM D0 vs. Th1 CM D5</i>	-13.24	-17.41 to -9.066	****	<0.0001
<i>Th2 CM D0 vs. Th1 EM D5</i>	-14.54	-18.71 to -10.36	****	<0.0001
<i>Th2 CM D0 vs. Th2 CM D5</i>	-2.207	-6.381 to 1.967	ns	0.2907
<i>Th2 CM D0 vs. Th2 EM D5</i>	-4.248	-8.422 to -0.07380	*	0.0463
<i>Th2 CM D0 vs. Th1 /17 CM D5</i>	-17.98	-22.16 to -13.81	****	<0.0001
<i>Th2 CM D0 vs. Th1 /17 EM D5</i>	-20.04	-24.21 to -15.87	****	<0.0001
<i>Th2 EM D0 vs. Th1 /17 CM D0</i>	-2.376	-6.550 to 1.798	ns	0.2559
<i>Th2 EM D0 vs. Th1 /17 EM D0</i>	-3.482	-7.656 to 0.6918	ns	0.0993
<i>Th2 EM D0 vs. Th1 CM D5</i>	-11.94	-16.12 to -7.771	****	<0.0001
<i>Th2 EM D0 vs. Th1 EM D5</i>	-13.24	-17.42 to -9.068	****	<0.0001
<i>Th2 EM D0 vs. Th2 CM D5</i>	-0.9120	-5.086 to 3.262	ns	0.6603
<i>Th2 EM D0 vs. Th2 EM D5</i>	-2.953	-7.127 to 1.221	ns	0.1600
<i>Th2 EM D0 vs. Th1 /17 CM D5</i>	-16.69	-20.86 to -12.51	****	<0.0001
<i>Th2 EM D0 vs. Th1 /17 EM D5</i>	-18.74	-22.92 to -14.57	****	<0.0001
<i>Th1 /17 CM D0 vs. Th1 /17 EM D0</i>	-1.106	-5.280 to 3.068	ns	0.5942
<i>Th1 /17 CM D0 vs. Th1 CM D5</i>	-9.569	-13.74 to -5.395	****	<0.0001
<i>Th1 /17 CM D0 vs. Th1 EM D5</i>	-10.87	-15.04 to -6.692	****	<0.0001
<i>Th1 /17 CM D0 vs. Th2 CM D5</i>	1.464	-2.710 to 5.638	ns	0.4815
<i>Th1 /17 CM D0 vs. Th2 EM D5</i>	-0.5767	-4.751 to 3.597	ns	0.7809
<i>Th1 /17 CM D0 vs. Th1 /17 CM D5</i>	-14.31	-18.49 to -10.14	****	<0.0001
<i>Th1 /17 CM D0 vs. Th1 /17 EM D5</i>	-16.37	-20.54 to -12.19	****	<0.0001
<i>Th1 /17 EM D0 vs. Th1 CM D5</i>	-8.462	-12.64 to -4.288	**	0.0002
<i>Th1 /17 EM D0 vs. Th1 EM D5</i>	-9.760	-13.93 to -5.586	****	<0.0001
<i>Th1 /17 EM D0 vs. Th2 CM D5</i>	2.570	-1.604 to 6.744	ns	0.2198
<i>Th1 /17 EM D0 vs. Th2 EM D5</i>	0.5296	-3.644 to 4.704	ns	0.7984
<i>Th1 /17 EM D0 vs. Th1 /17 CM D5</i>	-13.21	-17.38 to -9.031	****	<0.0001
<i>Th1 /17 EM D0 vs. Th1 /17 EM D5</i>	-15.26	-19.44 to -11.09	****	<0.0001
<i>Th1 CM D5 vs. Th1 EM D5</i>	-1.297	-5.471 to 2.877	ns	0.5325
<i>Th1 CM D5 vs. Th2 CM D5</i>	11.03	6.859 to 15.21	****	<0.0001
<i>Th1 CM D5 vs. Th2 EM D5</i>	8.992	4.818 to 13.17	***	0.0001
<i>Th1 CM D5 vs. Th1 /17 CM D5</i>	-4.743	-8.917 to -0.5690	*	0.0271
<i>Th1 CM D5 vs. Th1 /17 EM D5</i>	-6.799	-10.97 to -2.625	**	0.0022

<i>Th1 EM D5 vs. Th2 CM D5</i>	12.33	8.156 to 16.50	****	<0.0001
<i>Th1 EM D5 vs. Th2 EM D5</i>	10.29	6.115 to 14.46	****	<0.0001
<i>Th1 EM D5 vs. Th1 /17 CM D5</i>	-3.446	-7.620 to 0.7283	ns	0.1028
<i>Th1 EM D5 vs. Th1 /17 EM D5</i>	-5.502	-9.676 to -1.328	*	0.0112
<i>Th2 CM D5 vs. Th2 EM D5</i>	-2.041	-6.215 to 2.133	ns	0.3280
<i>Th2 CM D5 vs. Th1 /17 CM D5</i>	-15.78	-19.95 to -11.60	****	<0.0001
<i>Th2 CM D5 vs. Th1 /17 EM D5</i>	-17.83	-22.01 to -13.66	****	<0.0001
<i>Th2 EM D5 vs. Th1 /17 CM D5</i>	-13.74	-17.91 to -9.561	****	<0.0001
<i>Th2 EM D5 vs. Th1 /17 EM D5</i>	-15.79	-19.97 to -11.62	****	<0.0001
<i>Th1 /17 CM D5 vs. Th1 /17 EM D5</i>	-2.056	-6.231 to 2.118	ns	0.3244

**Table 3.26: Anova Fisher's LSD test on GATA3 expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells**

Fisher's LSD test	Mean Diff.	95.00% CI of diff.	Summary	Individual P Value
<i>Th1 CM D0 vs. Th1 EM D0</i>	0.3143	-6.481 to 7.109	ns	0.9247
<i>Th1 CM D0 vs. Th2 CM D0</i>	-10.75	-17.54 to -3.954	**	0.0033
<i>Th1 CM D0 vs. Th2 EM D0</i>	-13.76	-20.56 to -6.970	***	0.0003
<i>Th1 CM D0 vs. Th1 /17 CM D0</i>	0.7900	-6.005 to 7.585	ns	0.8124
<i>Th1 CM D0 vs. Th1 /17 EM D0</i>	-0.5392	-7.334 to 6.256	ns	0.8713
<i>Th1 CM D0 vs. Th1 CM D5</i>	-2.838	-9.632 to 3.957	ns	0.3973
<i>Th1 CM D0 vs. Th1 EM D5</i>	-7.389	-14.18 to -0.5938	*	0.0343
<i>Th1 CM D0 vs. Th2 CM D5</i>	-17.83	-24.63 to -11.04	****	<0.0001
<i>Th1 CM D0 vs. Th2 EM D5</i>	-13.67	-20.46 to -6.873	***	0.0004
<i>Th1 CM D0 vs. Th1 /17 CM D5</i>	-0.3405	-7.135 to 6.454	ns	0.9185
<i>Th1 CM D0 vs. Th1 /17 EM D5</i>	-1.176	-7.971 to 5.619	ns	0.7240
<i>Th1 EM D0 vs. Th2 CM D0</i>	-11.06	-17.86 to -4.269	**	0.0026
<i>Th1 EM D0 vs. Th2 EM D0</i>	-14.08	-20.87 to -7.284	***	0.0003
<i>Th1 EM D0 vs. Th1 /17 CM D0</i>	0.4757	-6.319 to 7.271	ns	0.8863
<i>Th1 EM D0 vs. Th1 /17 EM D0</i>	-0.8535	-7.648 to 5.941	ns	0.7977
<i>Th1 EM D0 vs. Th1 CM D5</i>	-3.152	-9.947 to 3.643	ns	0.3479
<i>Th1 EM D0 vs. Th1 EM D5</i>	-7.703	-14.50 to -0.9082	*	0.0279
<i>Th1 EM D0 vs. Th2 CM D5</i>	-18.15	-24.94 to -11.35	****	<0.0001
<i>Th1 EM D0 vs. Th2 EM D5</i>	-13.98	-20.78 to -7.187	***	0.0003
<i>Th1 EM D0 vs. Th1 /17 CM D5</i>	-0.6548	-7.450 to 6.140	ns	0.8440
<i>Th1 EM D0 vs. Th1 /17 EM D5</i>	-1.491	-8.285 to 5.304	ns	0.6548
<i>Th2 CM D0 vs. Th2 EM D0</i>	-3.015	-9.810 to 3.779	ns	0.3688
<i>Th2 CM D0 vs. Th1 /17 CM D0</i>	11.54	4.744 to 18.33	**	0.0018
<i>Th2 CM D0 vs. Th1 /17 EM D0</i>	10.21	3.415 to 17.01	**	0.0049
<i>Th2 CM D0 vs. Th1 CM D5</i>	7.912	1.117 to 14.71	*	0.0243
<i>Th2 CM D0 vs. Th1 EM D5</i>	3.361	-3.434 to 10.16	ns	0.3175
<i>Th2 CM D0 vs. Th2 CM D5</i>	-7.084	-13.88 to -0.2886	*	0.0417
<i>Th2 CM D0 vs. Th2 EM D5</i>	-2.918	-9.713 to 3.877	ns	0.3842
<i>Th2 CM D0 vs. Th1 /17 CM D5</i>	10.41	3.614 to 17.20	**	0.0042
<i>Th2 CM D0 vs. Th1 /17 EM D5</i>	9.573	2.778 to 16.37	**	0.0077
<i>Th2 EM D0 vs. Th1 /17 CM D0</i>	14.55	7.760 to 21.35	***	0.0002
<i>Th2 EM D0 vs. Th1 /17 EM D0</i>	13.23	6.431 to 20.02	***	0.0005
<i>Th2 EM D0 vs. Th1 CM D5</i>	10.93	4.132 to 17.72	**	0.0029
<i>Th2 EM D0 vs. Th1 EM D5</i>	6.376	-0.4189 to 13.17	ns	0.0646
<i>Th2 EM D0 vs. Th2 CM D5</i>	-4.068	-10.86 to 2.727	ns	0.2285
<i>Th2 EM D0 vs. Th2 EM D5</i>	0.09735	-6.698 to 6.892	ns	0.9767
<i>Th2 EM D0 vs. Th1 /17 CM D5</i>	13.42	6.629 to 20.22	***	0.0004
<i>Th2 EM D0 vs. Th1 /17 EM D5</i>	12.59	5.794 to 19.38	***	0.0008
<i>Th1 /17 CM D0 vs. Th1 /17 EM D0</i>	-1.329	-8.124 to 5.466	ns	0.6900
<i>Th1 /17 CM D0 vs. Th1 CM D5</i>	-3.628	-10.42 to 3.167	ns	0.2815
<i>Th1 /17 CM D0 vs. Th1 EM D5</i>	-8.179	-14.97 to -1.384	*	0.0204
<i>Th1 /17 CM D0 vs. Th2 CM D5</i>	-18.62	-25.42 to -11.83	****	<0.0001
<i>Th1 /17 CM D0 vs. Th2 EM D5</i>	-14.46	-21.25 to -7.663	***	0.0002
<i>Th1 /17 CM D0 vs. Th1 /17 CM D5</i>	-1.130	-7.925 to 5.664	ns	0.7343
<i>Th1 /17 CM D0 vs. Th1 /17 EM D5</i>	-1.966	-8.761 to 4.829	ns	0.5560
<i>Th1 /17 EM D0 vs. Th1 CM D5</i>	-2.298	-9.093 to 4.497	ns	0.4918
<i>Th1 /17 EM D0 vs. Th1 EM D5</i>	-6.850	-13.64 to -0.05463	*	0.0483
<i>Th1 /17 EM D0 vs. Th2 CM D5</i>	-17.29	-24.09 to -10.50	****	<0.0001
<i>Th1 /17 EM D0 vs. Th2 EM D5</i>	-13.13	-19.92 to -6.333	***	0.0005
<i>Th1 /17 EM D0 vs. Th1 /17 CM D5</i>	0.1987	-6.596 to 6.994	ns	0.9524
<i>Th1 /17 EM D0 vs. Th1 /17 EM D5</i>	-0.6370	-7.432 to 6.158	ns	0.8482
<i>Th1 CM D5 vs. Th1 EM D5</i>	-4.551	-11.35 to 2.244	ns	0.1796
<i>Th1 CM D5 vs. Th2 CM D5</i>	-15.00	-21.79 to -8.200	***	0.0001
<i>Th1 CM D5 vs. Th2 EM D5</i>	-10.83	-17.62 to -4.035	**	0.0031
<i>Th1 CM D5 vs. Th1 /17 CM D5</i>	2.497	-4.298 to 9.292	ns	0.4556
<i>Th1 CM D5 vs. Th1 /17 EM D5</i>	1.661	-5.134 to 8.456	ns	0.6184

<i>Th1 EM D5 vs. Th2 CM D5</i>	-10.44	-17.24 to -3.649	**	0.0041
<i>Th1 EM D5 vs. Th2 EM D5</i>	-6.279	-13.07 to 0.5162	ns	0.0685
<i>Th1 EM D5 vs. Th1 /17 CM D5</i>	7.048	0.2533 to 13.84	*	0.0427
<i>Th1 EM D5 vs. Th1 /17 EM D5</i>	6.213	-0.5824 to 13.01	ns	0.0713
<i>Th2 CM D5 vs. Th2 EM D5</i>	4.165	-2.629 to 10.96	ns	0.2179
<i>Th2 CM D5 vs. Th1 /17 CM D5</i>	17.49	10.70 to 24.29	****	<0.0001
<i>Th2 CM D5 vs. Th1 /17 EM D5</i>	16.66	9.862 to 23.45	****	<0.0001
<i>Th2 EM D5 vs. Th1 /17 CM D5</i>	13.33	6.532 to 20.12	***	0.0005
<i>Th2 EM D5 vs. Th1 /17 EM D5</i>	12.49	5.696 to 19.29	***	0.0009
<i>Th1 /17 CM D5 vs. Th1 /17 EM D5</i>	-0.8357	-7.631 to 5.959	ns	0.8018

**Table 3.27: Anova Fisher's LSD test on ROR $\gamma$ t expression in non-activated (D0) or activated (D5), Central Memory (CM) and Effector Memory (EM) subset of Th1, Th2 and Th/17 cells**

<b>Fisher's LSD test</b>	<b>Mean Diff.</b>	<b>95.00% CI of diff.</b>	<b>Summary</b>	<b>Individual P Value</b>
<i>Th1 CM D0 vs. Th1 EM D0</i>	0.8454	-7.592 to 9.282	ns	0.8379
<i>Th1 CM D0 vs. Th2 CM D0</i>	0.6227	-7.814 to 9.060	ns	0.8802
<i>Th1 CM D0 vs. Th2 EM D0</i>	0.2838	-8.153 to 8.721	ns	0.9452
<i>Th1 CM D0 vs. Th1 /17 CM D0</i>	-13.44	-21.87 to -5.000	**	0.0031
<i>Th1 CM D0 vs. Th1 /17 EM D0</i>	-30.25	-38.69 to -21.82	****	<0.0001
<i>Th1 CM D0 vs. Th1 CM D5</i>	-0.8691	-9.306 to 7.568	ns	0.8334
<i>Th1 CM D0 vs. Th1 EM D5</i>	-0.4158	-8.853 to 8.021	ns	0.9198
<i>Th1 CM D0 vs. Th2 CM D5</i>	-0.5469	-8.984 to 7.890	ns	0.8947
<i>Th1 CM D0 vs. Th2 EM D5</i>	-0.4940	-8.931 to 7.943	ns	0.9048
<i>Th1 CM D0 vs. Th1 /17 CM D5</i>	-18.11	-26.55 to -9.672	***	0.0002
<i>Th1 CM D0 vs. Th1 /17 EM D5</i>	-27.47	-35.90 to -19.03	****	<0.0001
<i>Th1 EM D0 vs. Th2 CM D0</i>	-0.2226	-8.660 to 8.214	ns	0.9570
<i>Th1 EM D0 vs. Th2 EM D0</i>	-0.5616	-8.999 to 7.875	ns	0.8919
<i>Th1 EM D0 vs. Th1 /17 CM D0</i>	-14.28	-22.72 to -5.845	**	0.0019
<i>Th1 EM D0 vs. Th1 /17 EM D0</i>	-31.10	-39.54 to -22.66	****	<0.0001
<i>Th1 EM D0 vs. Th1 CM D5</i>	-1.714	-10.15 to 6.723	ns	0.6787
<i>Th1 EM D0 vs. Th1 EM D5</i>	-1.261	-9.698 to 7.176	ns	0.7604
<i>Th1 EM D0 vs. Th2 CM D5</i>	-1.392	-9.829 to 7.045	ns	0.7364
<i>Th1 EM D0 vs. Th2 EM D5</i>	-1.339	-9.776 to 7.098	ns	0.7460
<i>Th1 EM D0 vs. Th1 /17 CM D5</i>	-18.95	-27.39 to -10.52	***	0.0001
<i>Th1 EM D0 vs. Th1 /17 EM D5</i>	-28.31	-36.75 to -19.88	****	<0.0001
<i>Th2 CM D0 vs. Th2 EM D0</i>	-0.3390	-8.776 to 8.098	ns	0.9346
<i>Th2 CM D0 vs. Th1 /17 CM D0</i>	-14.06	-22.50 to -5.622	**	0.0021
<i>Th2 CM D0 vs. Th1 /17 EM D0</i>	-30.88	-39.31 to -22.44	****	<0.0001
<i>Th2 CM D0 vs. Th1 CM D5</i>	-1.492	-9.929 to 6.945	ns	0.7184
<i>Th2 CM D0 vs. Th1 EM D5</i>	-1.039	-9.476 to 7.399	ns	0.8016
<i>Th2 CM D0 vs. Th2 CM D5</i>	-1.170	-9.607 to 7.267	ns	0.7772
<i>Th2 CM D0 vs. Th2 EM D5</i>	-1.117	-9.554 to 7.320	ns	0.7871
<i>Th2 CM D0 vs. Th1 /17 CM D5</i>	-18.73	-27.17 to -10.29	***	0.0001
<i>Th2 CM D0 vs. Th1 /17 EM D5</i>	-28.09	-36.53 to -19.65	****	<0.0001
<i>Th2 EM D0 vs. Th1 /17 CM D0</i>	-13.72	-22.16 to -5.283	**	0.0026
<i>Th2 EM D0 vs. Th1 /17 EM D0</i>	-30.54	-38.97 to -22.10	****	<0.0001
<i>Th2 EM D0 vs. Th1 CM D5</i>	-1.153	-9.590 to 7.284	ns	0.7803
<i>Th2 EM D0 vs. Th1 EM D5</i>	-0.6995	-9.137 to 7.738	ns	0.8656
<i>Th2 EM D0 vs. Th2 CM D5</i>	-0.8306	-9.268 to 7.606	ns	0.8407
<i>Th2 EM D0 vs. Th2 EM D5</i>	-0.7777	-9.215 to 7.659	ns	0.8507
<i>Th2 EM D0 vs. Th1 /17 CM D5</i>	-18.39	-26.83 to -9.956	***	0.0001
<i>Th2 EM D0 vs. Th1 /17 EM D5</i>	-27.75	-36.19 to -19.31	****	<0.0001
<i>Th1 /17 CM D0 vs. Th1 /17 EM D0</i>	-16.82	-25.25 to -8.379	***	0.0004
<i>Th1 /17 CM D0 vs. Th1 CM D5</i>	12.57	4.130 to 21.00	**	0.0052
<i>Th1 /17 CM D0 vs. Th1 EM D5</i>	13.02	4.584 to 21.46	**	0.0040
<i>Th1 /17 CM D0 vs. Th2 CM D5</i>	12.89	4.453 to 21.33	**	0.0043
<i>Th1 /17 CM D0 vs. Th2 EM D5</i>	12.94	4.506 to 21.38	**	0.0042
<i>Th1 /17 CM D0 vs. Th1 /17 CM D5</i>	-4.672	-13.11 to 3.765	ns	0.2643
<i>Th1 /17 CM D0 vs. Th1 /17 EM D5</i>	-14.03	-22.47 to -5.593	**	0.0022
<i>Th1 /17 EM D0 vs. Th1 CM D5</i>	29.38	20.95 to 37.82	****	<0.0001
<i>Th1 /17 EM D0 vs. Th1 EM D5</i>	29.84	21.40 to 38.27	****	<0.0001
<i>Th1 /17 EM D0 vs. Th2 CM D5</i>	29.71	21.27 to 38.14	****	<0.0001
<i>Th1 /17 EM D0 vs. Th2 EM D5</i>	29.76	21.32 to 38.20	****	<0.0001
<i>Th1 /17 EM D0 vs. Th1 /17 CM D5</i>	12.14	3.707 to 20.58	**	0.0067
<i>Th1 /17 EM D0 vs. Th1 /17 EM D5</i>	2.786	-5.651 to 11.22	ns	0.5020
<i>Th1 CM D5 vs. Th1 EM D5</i>	0.4533	-7.984 to 8.890	ns	0.9126
<i>Th1 CM D5 vs. Th2 CM D5</i>	0.3222	-8.115 to 8.759	ns	0.9378
<i>Th1 CM D5 vs. Th2 EM D5</i>	0.3751	-8.062 to 8.812	ns	0.9276
<i>Th1 CM D5 vs. Th1 /17 CM D5</i>	-17.24	-25.68 to -8.803	***	0.0003
<i>Th1 CM D5 vs. Th1 /17 EM D5</i>	-26.60	-35.03 to -18.16	****	<0.0001

<i>Th1 EM D5 vs. Th2 CM D5</i>	-0.1311	-8.568 to 8.306	ns	0.9747
<i>Th1 EM D5 vs. Th2 EM D5</i>	-0.07821	-8.515 to 8.359	ns	0.9849
<i>Th1 EM D5 vs. Th1 /17 CM D5</i>	-17.69	-26.13 to -9.256	***	0.0002
<i>Th1 EM D5 vs. Th1 /17 EM D5</i>	-27.05	-35.49 to -18.61	****	<0.0001
<i>Th2 CM D5 vs. Th2 EM D5</i>	0.05291	-8.384 to 8.490	ns	0.9898
<i>Th2 CM D5 vs. Th1 /17 CM D5</i>	-17.56	-26.00 to -9.125	***	0.0002
<i>Th2 CM D5 vs. Th1 /17 EM D5</i>	-26.92	-35.36 to -18.48	****	<0.0001
<i>Th2 EM D5 vs. Th1 /17 CM D5</i>	-17.61	-26.05 to -9.178	***	0.0002
<i>Th2 EM D5 vs. Th1 /17 EM D5</i>	-26.97	-35.41 to -18.54	****	<0.0001
<i>Th1 /17 CM D5 vs. Th1 /17 EM D5</i>	-9.358	-17.79 to -0.9208	*	0.0312

### 3.5 Discussion

In order to ascertain the function of ZEB2 in CD4<sup>+</sup> T cells, the first step was to determine exactly where it is expressed. In order to address this, a detailed and rigorous phenotyping analysis was carried out across the different subsets and lineages of human CD4<sup>+</sup> T cells. From these results I have determined in exactly which cell types ZEB2 is expressed and from this data I can infer a role for ZEB2. Our laboratory has previously shown that ZEB2 is repressed directly by FOXP3 in a feedforward loop mechanism (Brown et al., 2018) and others have shown that ZEB2 is directly induced by T-bet (Dominguez et al., 2015, Omilusik et al., 2015, van Helden et al., 2015).

The best way to characterise gene expression in populations of cells where flow cytometry is used, is to label a protein of interest with a fluorophore-conjugated antibody. To examine ZEB2 protein expression in human CD4<sup>+</sup> T cells, it requires intracellular labelling of the protein which is actually quite complicated and requires a good, robust antibody. By doing this, the protein can be detected directly in the cell populations, together with other cell surface antibodies used in separating various CD4<sup>+</sup> T cell populations and subsets, which is much more direct than quantitating ZEB2 mRNA expression in each population. However, this proved to be highly problematic, as described in Section 3.4.2, where I carried out a comprehensive set of troubleshooting experiments to detect intracellular ZEB2 protein. Even though I was able to easily detect ZEB2 intracellularly in HEK293T/17 cell lines overexpressing ZEB2, I could not get the antibody staining protocol to work in human primary T cells. I tried different cell fixation and permeabilization protocols, and different antibody sources at various concentrations. Despite much trouble shooting, I could not get the ZEB2 protein labelling to work in human CD4<sup>+</sup> T cells. As the time and effort on this aim was becoming unproductive, rather than spend time generating a new monoclonal antibody to human ZEB2, it was decided to end this approach. Clearly, the field will benefit from a better monoclonal antibody to ZEB2

that is sensitive and specific enough to detect the ZEB2 protein expression in human primary T cells.

My initial observations of ZEB2 expression using qRT-PCR in bulk Treg and Tconv revealed that ZEB2 was expressed much higher in a Tconv than a Treg. Consistent with this, low ZEB2 expression and chromatin accessibility in Naïve Treg and Tconv was observed in both Durek et al. (2016) and Calderon et al. (2019) respectively. These observations suggest that ZEB2 does not have a role in naïve CD4<sup>+</sup> T cells. However, my careful examination of ZEB2 expression across separated naïve, central memory, effector memory and effector memory RA<sup>+</sup> populations from Treg and Tconv revealed a more precise pattern of expression. Treg do express ZEB2, especially in the effector memory populations, but at a lower level compared with the Tconv populations (Figure 3.8). This expression pattern appears somewhat at odds with my overall results, but can be explained by the proportions of cells in the different populations. Treg Effector Memory (EM) only make up ~5-10% compared with the Central Memory (CM) Tregs that make up ~60-70% of cells in the peripheral blood. This, therefore, reduces the level of ZEB2 expression overall in bulk Treg. All these suggest a possible role of ZEB2 yet to be explored in the rare Treg EM. This is interesting and somewhat contradicts our previous finding that FOXP3 represses ZEB2, as in Treg EM both FOXP3 and ZEB2 expression is higher than in the naïve and CM subsets. Interestingly, studies from Mailloux and Epling-Burnette (2013) have shown that isolated Treg EM cells were significantly more suppressive than Treg CM cells *in vitro* so perhaps there is a role of ZEB2 in suppressive function which has yet to be explored.

CD4<sup>+</sup> T cells encounter a vast repertoire of immune challenges and in order to respond to these challenges quickly, specifically and effectively, naïve T cells reprogram their transcriptional landscape to enable a specific response to a particular immune challenge. The transcriptional landscape of these reprogrammed T cell subsets is coordinated by a “master transcription



factor” which induces the expression of an array of effector molecules and also specialised receptors, called chemokine receptors. Chemokine receptors allow the cell to migrate to the site of immune challenge. Different combinations of these receptors are expressed by specific T cell subsets and because they are expressed on the cell surface, we can use the expression patterns of these receptors to identify and isolate T cell subsets. Thinking that I would improve the efficiency of sorting the different T cell populations, in my initial experiments I used a pre-enrichment CD25 MACS step prior to chemokine receptor sorting. Upon examination of ZEB2 and master transcription factors expression level, inconsistency was observed between experiments. The reason is most likely owing to the CD25 pre-enrichment step where CD25<sup>hi</sup> activated cells might have been removed from the Tconv population, but included in the Treg population, resulting in overall ZEB2 expression across helper lineages of Tconv to be lower than their Treg counterpart. The explanation that most CD25<sup>hi</sup> CD127<sup>+</sup> Tconv population express higher levels of ZEB2 is consistent with previous studies in mice overexpressing ZEB2 showing that its CD4/8 double negative populations had increased CD25 protein (Goossens et al., 2019). Interestingly, RNA-seq data of various Treg and Tconv helper lineage populations from Höllbacher et al. (2020) has also shown that there was no ZEB2 enrichment in any populations. In their studies, Höllbacher et al. (2020) carried out CD25 MACS pre-enrichment as well, potentially also biasing CD25<sup>lo</sup> Tconv in their helper lineage population sort. These observations and studies from other groups suggested the need to replace CD25 MACS pre-enrichment in favour of a CD45RA based pre-enrichment steps. Previous experiments have shown that naïve Treg and Tconv have the lowest expression of ZEB2 and these cell populations were gated out during chemokine receptor sorts. Therefore, CD45RA pre-enrichment retains CD25<sup>hi</sup> cells to better represent the helper lineage subsets during the sort and increased sort efficiency.

As my first major finding after using CD45RA MACS pre-enrichment, I demonstrated that ZEB2 expression is almost exclusively found in Th1 cells. This observation was further supported by RNA-seq data mined from Schmiedel et al. (2018), where ZEB2 was also exclusively found in Th1. This raised the possibility that ZEB2 expression is controlled by the Th1 lineage defining transcription factor T-bet, as has been proposed in other cells in mouse Omilusik et al. (2015), Dominguez et al. (2015) and van Helden et al. (2015) showed that T-bet directly drives expression of ZEB2 in mouse CD8<sup>+</sup> T cells and NK cells, and in Figure 3.14 the expression profile shows that T-bet and ZEB2 are expressed predominantly in the same Th1 CD4 population, and this supports the likely co-regulation of ZEB2 by T-bet in human. However, the observation that ZEB2 is not expressed robustly in the strongly T-bet positive Th1/17 cells, suggests it is not a simple binary T-bet controlled, transcriptional regulation of ZEB2, and that it is possible that other mechanisms, possibly another transcription factor or microRNAs, inhibit ZEB2 expression in Th1/17 cells. Interestingly, the high expression of miR-155 in both the mouse and human Th17 (closely related to Th1/17) cells relative to Th1, Th2 or iTreg were reported by Escobar et al. (2014) suggesting possible role for miR155 in the Th1/17 helper lineage population. As mentioned above, our previous published data by Brown et al. (2018) showed that ZEB2 is post transcriptionally regulated by miR-155 and therefore even if T-bet is highly expressed in Th1/17 cells, the presence of miR-155 might predominate over T-bet and repress ZEB2. We have shown that a feedforward regulatory mechanism involving FOXP3 and FOXP3 induced miR-155 suppresses ZEB2 in Treg (Figure 1.8). From this data we can infer that in Th1 Treg and Th17 Treg subsets, ZEB2 is most likely suppressed by this FOXP3 and miR-155 feedforward mechanism and thus FOXP3 predominates over T-bet to prevent ZEB2 activation in all Treg. However, this feedforward mechanism may not apply in the Effector Memory Treg Th1 lineage populations Figure 3.8.

The second major findings that ZEB2 may also be induced by activation and not entirely dependent on T-bet expression agree in part with the studies of Serroukh et al. (2018). In their studies, Serroukh et al observed inconsistent ZEB2 expression in in-vitro differentiated effector T cells (Th1, Th2 & Th17 polarised cells) at different time points post-activation, similar to our findings (Figure 3.20). As T-bet is known to induce ZEB2 expression (Dominguez et al., 2015, Omilusik et al., 2015, van Helden et al., 2015), it was surprising that ZEB2 also comes on in activated Th2 and Th1/17 CM and EM even though T-bet does not. This could suggest that during activation something else is turning on ZEB2 in the Th2 or something that normally represses ZEB2 has been turned off. Therefore, ZEB2 expression may not always be dependent on T-bet expression. It was also very interesting that GATA3 is still pretty restricted to Th2 although it does come on in Th1 somewhat, during activation confirming findings from Lantelme et al. (2001) that there are low basal levels of GATA3 expression in Th1 cells upon activation. Studies from Ho et al. (2009) and Wang et al. (2013) suggest that GATA3 is not only a master transcription factor of Th2, but is also important in controlling T cell maintenance and proliferation, downstream of TCR and cytokine signals which may explain the basal level of GATA3 expression observed in other helper lineage population subsets. In contrast, ROR $\gamma$ t expression does not change at all between non-activated and activated cells. This suggests that ROR $\gamma$ t is completely restricted to Th1/17. Overall, ZEB2 expression may be required in a Th1 EM in the periphery owing to its unique expression, however during *in vitro* stimulation CM and EM acquire ZEB2 expression, suggesting that ZEB2 may be involved in activation and that Th1 EM cells expressing ZEB2 may be in a semi activated state and ready to carry out its effector function. Alternatively, expression of ZEB2 in the CD4<sup>+</sup> T cell populations activated with IL-2 and anti-CD3/CD28 may not fully reflect the environmental milieu encountered by T cells in the human periphery and thus may be somewhat artifactual. Therefore, it will be interesting to examine if these findings are consistent *in vivo* in mice or in human samples with

exposure to a Th1, Th2 or Th17 mediated response. This poses the question will ZEB2 remain exclusively in the T-bet+ CD4+ T cells during infection or, more broadly, consistent with my activation studies (Figure 3.20)? As such experiments have not been conducted *in vivo*, our data represent new insights into this topic, with the caveat that more work would be required to determine whether or not this is simply a feature of invitro activation, and that under more physiological conditions , this does not occur.

Overall, the results of this chapter support the hypothesis that ZEB2 is predominantly expressed in the effector memory and T-bet expressing Th1 Effector Memory subset of human CD4+ T cells. However, new light has been shed on the direct regulation of ZEB2 by T-bet, as T-bet expression does not always correlate with ZEB2 expression and hence warrants further investigation into other mechanisms involved in inhibiting ZEB2. Additionally, ZEB2 was also not exclusively expressed in Th1 EM post activation, suggesting a potential activation component in ZEB2 expression in all of the Tconv helper lineage populations. These results indicate the importance of ZEB2 in Th1 Effector Memory cells, especially in the periphery in which the cells are in a semi activated state ready to carry out effector function when encountering pathogens.

## **CHAPTER 4: ZEB2 FUNCTION IN TH1 EFFECTOR MEMORY CELLS**

## 4.1 Introduction

In the previous chapter, I determined that ZEB2 was found to be expressed almost exclusively in human Th1 Effector Memory (EM) cells where T-bet was also expressed. However, not all T-bet<sup>+</sup> cells in the CD4<sup>+</sup> T cell expresses high ZEB2. Interestingly, upregulation of ZEB2 expression was also found in other Tconv helper populations after in- vitro activation, suggesting an activation component resulting in ZEB2 induction. Although my finding that ZEB2 is almost completely confined to the Th1 EM compartment in peripheral blood, suggesting a high likelihood that it will be involved in the effector function of Tconv cells, further analysis is required to fully elucidate its role. Previous studies have shown that ZEB2 plays a role in effector function and terminal differentiation of mouse CD8<sup>+</sup> T cell (Dominguez et al., 2015, Omilusik et al., 2015) and CTL (Patil et al., 2018, Serroukh et al., 2018) as well as migration and self-renewal in metastatic cancer (Brown et al., 2018, Goossens et al., 2015, Li et al., 2016a), and therefore ZEB2 may potentially function quite similarly in a Th1 EM cell. Since CD8 CTL are very different cell types compared with Th1 EM; other than secreting effector cytokines such as IFN $\gamma$ , CD8 CTL also secrete cytotoxic molecules to carry out their killing activities, this further suggests that the role of ZEB2 in the differentiation and effector function of Th1 EM may be uniquely different and novel compared with other cell types.

A key question therefore, to be addressed in this chapter is, what is ZEB2 regulating when expressed exclusively in the Th1 EM cells? Before trying to determine the role of ZEB2 in Th1 EM, a further insight into the formation and function of Th1 effector memory is required. When 90% of the Th1 effector cells die during the contraction phase after clearance of pathogens this leaves a residual population of long-lived cells. These cells, which are known as memory cells, are predominantly quiescent with self-renewal capabilities and long-term survival (Geginat et al., 2014). Th1 memory cells are heterogeneous, however, and are proposed to exist in at least two classes. Th1 EM express homing receptors that facilitate migration to non-lymphoid sites

of inflammation and produce IFN $\gamma$  within several hours of TCR stimulation and this is the class of Th1 memory cells that express high ZEB2. Whereas Th1 CM, (which express low ZEB2 as presented in Chapter 3), do not produce any of the effector cytokines immediately after stimulation through the TCR, although they do secrete IL-2 and proliferate extensively and can acquire effector cytokine production later. It is possible that ZEB2 allows Th1 EM to circulate through these non-lymphoid sites of inflammation which would allow for secondary effector responses when encountering recurring intracellular pathogens.

Here in this chapter, I will focus on defining the role of ZEB2 in Th1 EM cells. ZEB2 is a known transcription factor that orchestrates changes to global gene expression (Li et al., 2016b). Therefore by manipulation of ZEB2 gene expression, by either gene ablation or enforced expression, this allows for a detailed analysis of changes to cell phenotype and genotype mediated by loss or gain of gene function. To identify potential targets of ZEB2 in Th1 EM, I used genome editing tools that deleted endogenous ZEB2 from primary human Th1 EM cells using CRISPR/Cas9 technology. An RNA-seq approach was then used to capture the global transcriptome profile differences between edited and non-edited cells using deep-sequencing technologies to elucidate the role of ZEB2 in Th1 EM. To obtain robust and consistent gene knockout and with cell numbers high enough for downstream applications required extensive optimisation. In this chapter, I explain the design of the sgRNAs sequence and the two different CRISPR/Cas9 delivery strategies to knockout ZEB2. I first started with a single sgRNA delivery approach, as demonstrated by Schumann et al. (2015), followed by a single cell cloning protocol to isolate clones with full knockout. This approach was time consuming owing to low editing efficiency in a cell pool and long-term culture to expand single cell clones. Additionally, because the long-term culture could easily result in aberrant findings, I then tested a second method using a triple sgRNAs delivery approach, as demonstrated by Seki and Rutz (2018), that yielded editing efficiencies high enough so that cells could be used as a pool for

downstream experiments. To further understand the role of ZEB2 in the CD4<sup>+</sup> T cells compartment, I used lentivirus to deliver ZEB2 into naïve Tconv and naïve Treg which normally express low ZEB2 as indicated in the previous chapter. In a scenario where these cells overexpress ZEB2, I speculated as to a potential gain of function in a cell where its expression is usually low? Perhaps gaining Th1 function?

This chapter demonstrates that ablation of ZEB2 results in loss of effector function but gain of cytotoxic response and upregulation of other chemokine receptors. In addition, ZEB2 deleted Th1 EM cells would most likely have reduced survival during environmental stress, reduced migratory capabilities by calcium ion homeostasis.



## 4.2 Aims and Hypothesis

The hypothesis for this chapter is that ZEB2 is required to maintain Th1 function and fidelity

The focus of this chapter is to address Aim 2 of my PhD proposal.

- 2.1. To design sgRNAs target sequences for functional knockout of ZEB2.
- 2.2. To delete endogenous ZEB2 in human primary Th1 EM cells using an optimised protocol of CRISPR/Cas9 delivery.
- 2.3. To identify global transcriptome profile changes caused by the loss of ZEB2 from RNA-seq differential expression data analysis between edited (knockout) and non-edited (wildtype) Th1 EM cells.
- 2.4. To determine gene expression changes by overexpression of ZEB2 using a lentivirus delivery system in Naïve Treg and Tconv.

## 4.3 Material & Methods

### 4.3.1 Cell Isolation

Buffy coats of healthy donors were obtained from the Australian Red Cross with written informed consent and approval by the WCHN Human Research Ethics Committee under REC1596/08/2019 for research purposes. CD4<sup>+</sup> T cells were first enriched using RosetteSep™ Human CD4<sup>+</sup> T cells enrichment cocktail (STEMCELL Technologies) as indicated in Section 2.1.2 followed by CD45RA MACS to enrich for CD45RA<sup>-</sup> memory CD4 for Th1 EM cell isolation or CD45RA<sup>+</sup> naïve CD4 for naïve Treg and naïve Tconv isolation.

For deletion of ZEB2 in Th1 EM cells (single sgRNAs approach) or Th1 cells (triple sgRNAs approach), enriched memory CD4<sup>+</sup> T cells were rested overnight in resting medium (CX-VIVO) and incubated at 37°C with 5% CO<sub>2</sub>. The next day cells were then labelled with monoclonal surface antibodies to CD4, CD25, CD45RA, CD62L, CXCR3, CCR6, CCR4 and CCR10. These markers were selected based on antibodies described in Table 3.4 which was used in previous chapters to identify Th1 EM cells or Th1 for flow cytometry following the gating strategy shown in Figure 3.12 and Figure 3.15.

For overexpression of ZEB2 in both naïve Treg and Tconv experiment, enriched naïve CD4<sup>+</sup> T cells were rested overnight in resting medium (CX-VIVO) and incubated at 37°C with 5% CO<sub>2</sub>. The next day cells were then labelled with monoclonal surface antibodies to CD4, CD25, CD45RA, CD62L. These markers were as used in previous chapters to identify naïve Treg (CD4<sup>+</sup>, CD25<sup>hi</sup>, CD127<sup>lo</sup>, CD45RA<sup>+</sup>, CD62<sup>+</sup>) and naïve Tconv (CD4<sup>+</sup>, CD25<sup>lo</sup>, CD127<sup>hi</sup>, CD45RA<sup>+</sup>) cells using flow cytometry following the gating strategy shown in Figure 3.7 (antibodies described in Table 3.2).

Surface labelled cells were then sort purified with a FACS Fusion flow cytometer (BD Biosciences, San Jose, USA) (Section 2.1.6). Sort purified Th1 EM cells or naïve Treg and

naïve Tconv cells were washed, pelleted and resuspended in T cell culture medium (CX-VIVO + 100 units/mL IL-2). Cells were checked for viability and counted as described above (Section 2.1.1) before incubation at 37°C with 5% CO<sub>2</sub> for further use.

### 4.3.2 CRISPR/Cas9 Deletion of ZEB2

#### Preparation of cells

Since I have already identified that ZEB2 is expressed highly in the Th1 EM cells and Th1 cells, these cells will be used for endogenous ZEB2 deletion experiments. To generate the cell numbers required for deletion using the single sgRNA approach, isolated Th1 or Th1 EM cells, as described above, were cultured at 1x10<sup>6</sup> cells/ml in T cell activation culture medium (CX-VIVO + 100 units/mL IL-2) and expanded with 1:1 anti-CD3/CD28 magnetic beads (Dynabeads® Human T-Activator CD3/CD28, Life Technologies) for 2 weeks to achieve 2-4x10<sup>6</sup> cells before transfection. To prepare the cells for deletion using the triple sgRNA approach, isolated Th1 or Th1 EM as described above, were cultured at 1x10<sup>6</sup> cells/ml in T cell activation culture medium (CX-VIVO + 100 units/mL IL-2) and activated with 1:1 anti-CD3/CD28 magnetic beads (Dynabeads® Human T-Activator CD3/CD28, Life Technologies) for 3 days prior to transfection.

#### Pre-complexing sgRNA-Cas9 to form Ribonucleoprotein (RNP)

The sgRNA sequences were designed as described in Section 2.3.1. In this chapter two approaches were used to carry out ZEB2 knockout experiments. Initially, one sgRNA was used (ZEB2 sgRNA #1: GGUGAACUAUGACAAUGUAG, Table 2.4 Synthego, California, USA), but this was further refined for subsequent experiments where three sgRNAs were used (ZEB2 sgRNAs #1: GGUGAACUAUGACAAUGUAG, #2: UAUGACAAUGUAGUGGACAC and #3: CACAGGUUCUGAAACAGAUG, Table 2.4, Synthego, California, USA), since this improved knockout efficiency. First, the sgRNAs were mixed with Cas9 nuclease to form RNP complexes. For each nucleofection (using 1x10<sup>6</sup>-2x10<sup>6</sup> cells), of chemically modified sgRNA

targeting ZEB2 or non-targeting control (6 $\mu$ L equal to 240pmol, Synthego, California, USA), 2NLS Cas9 nuclease (4 $\mu$ L equal to 80pmol, Synthego, California, USA) and Nucleofector™ solution (2 $\mu$ L, Human T Cells Nucleofector™ Solution, Lonza, Switzerland) were combined in a PCR tube and gently mixed by pipetting. The reaction mixes were then incubated at room temperature for at least 10 minutes to form the RNP complex.

### Nucleofection

To allow the RNP complex to enter the cells, an electroporation-based transfection method known as Nucleofection™ (Lonza trademarked) was used. 2mL of T cell culture medium (CX-VIVO + 100 units/mL IL-2) per well of a 24-well plate was pre-warmed in an incubator for 45 minutes at 37°C with 5% CO<sub>2</sub>. Anti-CD3/CD28 magnetic beads were removed from Th1 EM cells (see methods for beads removal in Section 2.1.7) after 1-2 weeks expansion as mentioned above, prior to Nucleofection™ to prevent the magnetic beads from interfering with the electric pulse during Nucleofection™ which may harm the cells. 1x10<sup>6</sup>-2x10<sup>6</sup> cells were then resuspended in 100 $\mu$ L of Nucleofection™ solution (pre-warmed to room temperature, Human T Cells Nucleofector™ Solution, Lonza, Cat# VPA-1002, Switzerland). The T cells were mixed gently and incubated with 12 $\mu$ L RNP at room temperature for 2 minutes. The cell/RNP mix was transferred to Nucleofection™ cuvette (Human T Cell Nucleofector™ Kit, Lonza, Cat# VPA-1002, Switzerland). Cells were electroporated using a Nucleofector™ 2b Device (Cat#AAB-1001, Lonza, Switzerland) using the T-020 pulse setting for activated T cells. After nucleofection, 500 $\mu$ L of prewarmed T cell medium was slowly added to the cuvette and a transfer pipette was used to transfer the nucleofected cells into 24-well plates. Nucleofected cells were incubated at 37°C overnight and the following day half of the culture medium was replaced with fresh T cell culture medium. In experiments where single sgRNAs were used for nucleofection, cells were then single cell sorted and cultured for expansion of cell clones (see method in Section 2.3.3). In experiments where triple sgRNAs were used for nucleofection, 48

hours post-nucleofection, live cells were sorted based on forward versus side scatter, genomic DNA was isolated and then RNA was isolated as described in Chapter 2. The same downstream process was also carried out on expanded clones.

#### Determination of knockout efficiency

Genomic DNA was extracted according to methods in Section 2.3.4, and PCR primers (ZEB2 PCR Primer F: TTTCCTGACATGGTTGAGTAATTC, ZEB2 PCR Primer R: AATCTCGTTGTTGTGCCAGG, Integrated DNA Technologies, Inc., USA) were used to amplify the region flanking the CRISPR target sites, as shown in Figure 2.3. Following PCR amplification, the PCR product was then electrophoresed on a 1% Agarose gel at 80V for 1 hour to verify the amplification of a single band of the correct size. A diffuse band of approximately the correct size, suggesting multiple fragment sizes, indicates that CRISPR/Cas9 has most likely successfully targeted the ZEB2. The diffuse band indicates potential insertions or deletions in the amplified target region. Sanger sequencing was then carried out on ZEB2 KO and WT DNA samples from the Nucleofected™ T cells to determine the presence of insertions and deletions (INDELS) in the KO (edited) genomic DNA compared with the WT (unedited) genomic DNA and thus calculate the knockout efficiency in the ZEB2 KO (edited) T cells. To prepare the DNA for sequencing, the PCR amplicon was first purified using the NucleoSpin™ Gel and PCR Clean-up Kit (Cat#740609, Macherey-Nagel™, Düren, Germany) as per the manufacturers' protocols. Sequencing reactions were then set up following the AGRF Sanger sequencing sample submission guidelines, as indicated in Section 2.3.5. ICE analysis software (Synthego, USA) was then used to carry out a quantitative assessment of genome editing using the raw data output from the Sanger sequencing. The software compared the sequence traces of amplicons generated from genomic DNA isolated from both the edited and unedited pools of cells as detailed in Section 2.3.6.

### 4.3.3 RNA-seq

#### Library preparation

In order to look for global transcriptomic profile changes after deleting endogenous ZEB2 in Th1 EM, RNA-sequencing was carried out. Firstly, cDNA libraries were generated from isolated Th1 EM WT and KO RNA with a RIN score > 7 using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) (as described in Section 2.4.1). All procedures were performed according to the protocols suggested by the manufacturers as indicated in Section 2.4.3. Once the individual cDNA libraries were generated, their quality could be assessed using an Automated Electrophoresis System (Experion™ Automated Electrophoresis System, Bio-Rad Laboratories, Inc., USA) using the Experion DNA 1K Analysis Kit (Bio-Rad Laboratories, Inc., USA). Quantification of individual libraries was performed in a real-time set-up with KAPA Library Quantification Kits (Cat#07960140001, Roche, Basel, Switzerland). Individual libraries with different index barcodes were pooled for multiplexing and processed with the HiSeq 2500 System (Illumina, USA).

#### Statistical analysis

Alignment-free quantification methods, Salmon (version 1.2.1), estimated the count data based on indexes built from the human transcriptome (GRCh38). All the quantification methods were run at both the gene and transcript levels as detailed in Section 2.4.4 and 2.4.5. All of the statistical analysis (Section 2.4.6) and plots were carried out using R environment package (version 4.0). Scripts used in RNA-seq analysis can be found in the appendix.

### 4.3.4 Lentivirus overexpression of ZEB2

#### Preparation of cells

In the previous chapter, I have identified that ZEB2 expression was quite low in both naïve Treg and naïve Tconv. I was interested in examining the effects of ZEB2 when introduced into cells normally expressing low ZEB2. Therefore, in these experiments, ZEB2 overexpressing

lentivirus (Figure 2.8) was used to deliver ZEB2 into both naïve Treg and naïve Tconv. Naïve Treg and naïve Tconv cells were isolated as described above and cultured at  $1 \times 10^6$  cells/ml in T cell culture medium (CX-VIVO + 100 units/mL IL-2) overnight.

#### Lentivirus transduction of T cells

The next day, 3:1 anti-CD3/CD28 magnetic beads (Dynabeads® Human T-Activator CD3/CD28, Life Technologies) were added for 3 hours prior to lentivirus transduction. After activation, cells were transduced by adding the ZEB2 overexpressing lentivirus (LV411-ZEB2-IRES-hrGFP, packaged in Section 2.5.5) in the presence of Polybrene ( $8 \mu\text{g/mL}$ ) at a Multiplicity of Infection (MOI) of 10-20 ratio of virus: cells and then incubated at  $37^\circ\text{C}$  with 5%  $\text{CO}_2$  overnight. The next day, half of the medium was replaced with fresh medium without Polybrene. Transduction efficiency was determined 24 hours and 48 hours post transduction by flow cytometric analysis of % GFP+ cells as described above (Section 2.1.6 & Section 2.5.6). Transduced cells were further expanded to achieve required cell numbers (Section 2.1.7) before lysing cells (Qiazol™, QIAGEN, Germany) for RNA.

#### **4.3.5 Quantitative Real Time PCR**

Total RNA was isolated using Qiagen miRNeasy® Mini kit following the manufacturer's protocol. For mRNA quantitation, cDNA was generated from total RNA using the Qiagen QuantiTect kit as detailed in Section 2.2.3. KAPA SYBR® FAST qPCR kit was used for the subsequent quantitative real-time PCR following methods in Section 2.2.4. For CRISPR/Cas9 ZEB2 deletion using single sgRNAs, expression of S1PR5, EPCAM and IFNG were measured in WT and KO Th1 EM clones. For CRISPR/Cas9 ZEB2 deletion using triple sgRNAs, expression of ZEB2, T-bet and IFNG were measured in WT and KO Th1 EM cells pools. For lentivirus overexpression of ZEB2, expression of ZEB2, T-bet, FOXP3 and IL-10 were measured in ZEB2 transduced or untransduced naïve Treg and naïve Tconv. All qRT-PCR primer sequences used in gene expression measurement are as detailed in

Table 2.3.

### 4.3.6 Th1 polarisation

In the previous chapter, I determined that ZEB2 was mostly expressed in memory Tconv compared with naïve Tconv. Hence, in the next this experiments I examined whether ZEB2 has a role in promoting the differentiation of naïve CD4 to Th1 cells, under Th1 polarising condition, or if it is only expressed after differentiation of cells. To investigate this, naïve Tconv were isolated from enriched CD4<sup>+</sup> T cells using EasySep™ Human Naïve CD4<sup>+</sup> T Cell Isolation Kit (STEMCELL Technologies). Isolated naïve Tconv cells were activated using 1:1 CD3 and CD28 antibody coated magnetic beads (Dynabeads® Human T-Activator CD3/CD28, Life Technologies) and cultured in T cell culture medium (CX-VIVO + 100 units/mL IL-2) for 3 days before Nucleofection™ using the triple sgRNA approach to delete endogenous ZEB2 (see methods in Section 4.3.2). After Nucleofection™, cells were cultured in T cell culture medium (CX-VIVO + 100 units/mL IL-2) containing either Th1 polarisation reagents from the Human Th1 Cell Differentiation Kit (R&D Systems, US) or culture medium alone for 5 days at 37°C following the manufacturer's protocol (as described in Section 2.1.8. At day 5, the cells were removed from the antibody-coated magnetic beads and expression of the Th1 defining cytokine, IFN $\gamma$ , was measured to assess the proportion of cells that had polarised from naïve CD4<sup>+</sup> T cells to Th1 cells (see intracellular cytokine staining methods in Section 2.1.9).



## 4.4 Results

### 4.4.1 Optimising CRISPR/Cas9 techniques to knockout ZEB2 in Tconv Th1 EM

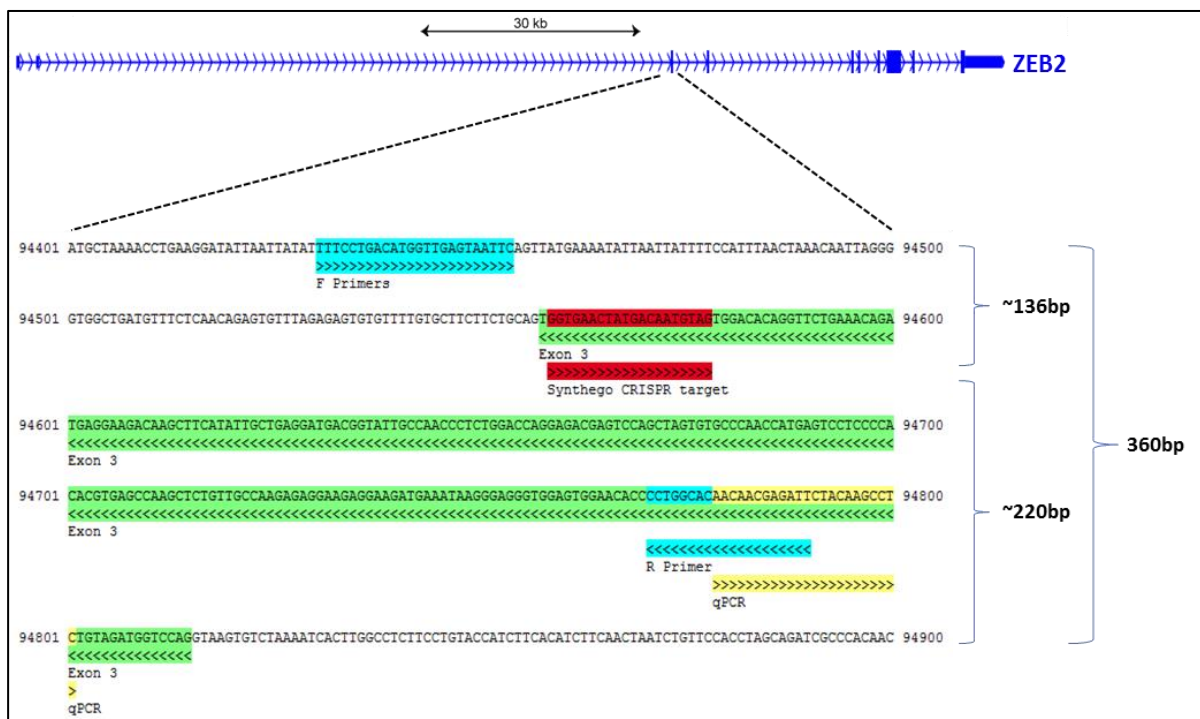
Having shown that ZEB2 is confined to the Th1 EM cell population, I endeavoured to define its role in these cells by examining the consequences of deleting it in freshly isolated human Th1 EM cells. This was carried out using CRISPR/Cas9 to delete the ZEB2 gene. Two different approaches were tested for deleting ZEB2 from Th1 EM cells: single sgRNAs or triple sgRNAs. These sgRNAs guides were designed to target exon 3 of the ZEB2 gene (as described in Section 2.3.1) as targeting this region was predicted to disrupt the translation of the ZEB2 protein. Forward and reverse PCR primers were also designed to flank the CRISPR target as shown in Figure 4.1 to verify that it is targeting the correct region. Initially, a system using one sgRNA was tested (ZEB2 sgRNA #1, indicated in Table 2.4), as this method has been reported to achieve successful gene editing with frequencies between 50% and 90% when transfecting activated human primary T cells using CRISPR/Cas9 RNP complexes (Schumann et al., 2015, Hendel et al., 2015, Rupp et al., 2017). However, owing to the low efficiency of this approach using single sgRNA, clonal expansion was required to purify the ZEB2 deleted Th1 EM clones. In subsequent experiments, this approach was refined using three sgRNAs (ZEB2 sgRNAs #1, #2 and #3, indicated in Table 2.4) as demonstrated by Seki and Rutz (2018) since this improves knockout efficiency. Such optimisation and improvement allowed downstream analysis to be carried out on ZEB2 deleted pools and therefore omitted the time-consuming clonal expansion steps.

### 4.4.2 A single sgRNA CRISPR/Cas9 deletion of ZEB2

#### Knockout efficiency verification

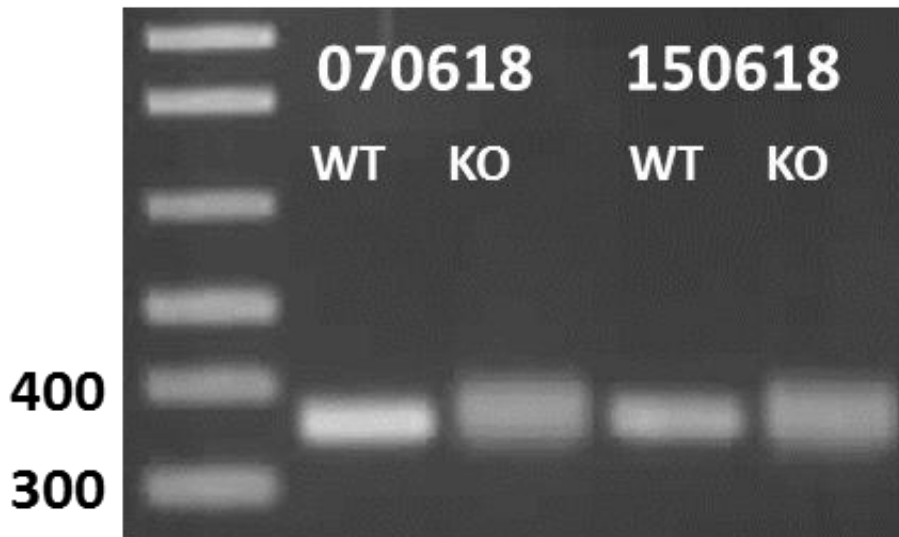
As mentioned above, initial ZEB2 knockout experiments were carried out using a single sgRNA guide system. Th1 EM cells were first isolated, and activated for 3 days before nucleofection with either ZEB2 targeting or non-targeting CRISPR/Cas9 RNPs. Cells were then rested for 48

hours to allow for gene editing to occur (see methods in Section 4.3.1 and Section 4.3.2). Before carrying out any downstream RNA-seq analysis on cells with ZEB2 knockout (KO) compared with Control (WT) cells, I first ascertained how efficient the gene knockout in the Th1 EM cells was. Genomic DNA was isolated from ZEB2 KO and WT cells and PCR amplification of the predicted area of deletion was carried out using PCR primers spanning the targeted area of the gene (Figure 4.1) (protocol in Section 2.3.4). After PCR amplification, the amplicons were then electrophoresed (80V) by agarose gel (1 %) electrophoresis, as described in Section 4.3.2. A band of 360bp was easily visible (Figure 4.2) indicating the amplified region and in knockout samples (samples #070618 and #150618) this band was diffuse compared with WT samples. This diffuse band is indicative of different sized DNA amplicons and suggestive of Insertions or Deletions (INDELs) generated by the CRISPR/Cas9 system and indicating that gene editing has occurred.



**Figure 4.1: Schematic of single CRISPR guide targeting ZEB2 exon 3**

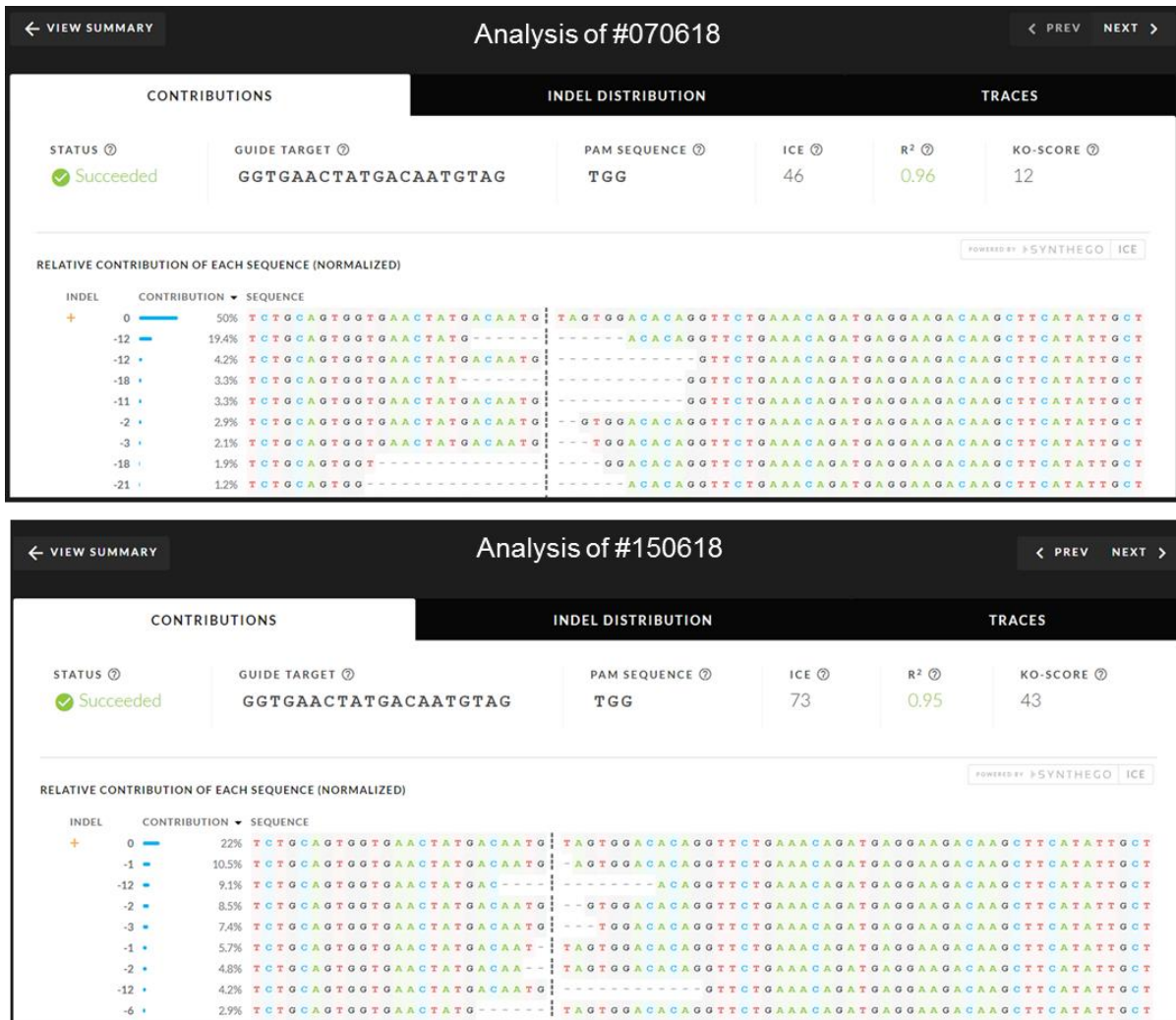
CRISPR guide target (red) designed by Synthego to target exon 3 (green) of ZEB2 gene flanked by forward and reverse primers (blue).



**Figure 4.2: Gel image of genomic PCR post CRISPR/Cas9 using single sgRNA**

Post genomic PCR of samples with either WT or KO from 2 donors were electrophoresed using 1% Agarose and 80V.

However, in order to determine the editing efficiency, analysis was carried out using a prediction tool called Inference of CRISPR Edits (ICE, Synthego, USA) (described in Section 2.3.6). Analysis using the ICE tool required Sanger sequencing of the PCR amplicons generated from the genomic DNA target region isolated from the ZEB2 KO and WT cells. The ICE analysis returns a dataset of INDELS and a KO score for the sequences detected, based on the calculation that a frameshift mutation would result in a null or non-functional protein. This tool additionally considers that a 21bp deletion as likely to result in loss of protein owing to a large deleted region, regardless of frameshift. Similarity in INDEL and KO scores, suggests efficient targeting of the guides. Differences in these scores, whereby the INDEL score is higher than the KO score, suggests that while a sgRNA guide may have targeted the gene, productive editing did not take place. There was much variation in knockout scores between donors and frequently a discrepancy between INDEL and KO score: for instance, donor #070618 had an editing efficiency of 46% but with only 12% resulting in gene deletion. In contrast, donor #150618 had an editing efficiency of 73% where 43% of those would result in knockouts (Figure 4.3).



**Figure 4.3: ICE analysis of single sgRNAs ZEB2 KO Th1 EM prior to single cell cloning**  
 The guide sequence, Protospacer Adjacent Motif (PAM; CRISPR/Cas9 cleavage site), ICE Score, R<sup>2</sup>, and Knockout Score are all shown for a ZEB2 edited sample for donor #070618 (top) and donor #150618 (bottom). Below this information, the inferred sequences present in the edited KO population (“Sequences”) and their relative representation in the edited KO pool (“Contribution”) are shown. For each sequence, the number of nucleotides inserted (+) or deleted (-) was indicated under “INDEL”. The black vertical dotted line represents the cut site, and “+” symbol on the far left marks the wildtype.

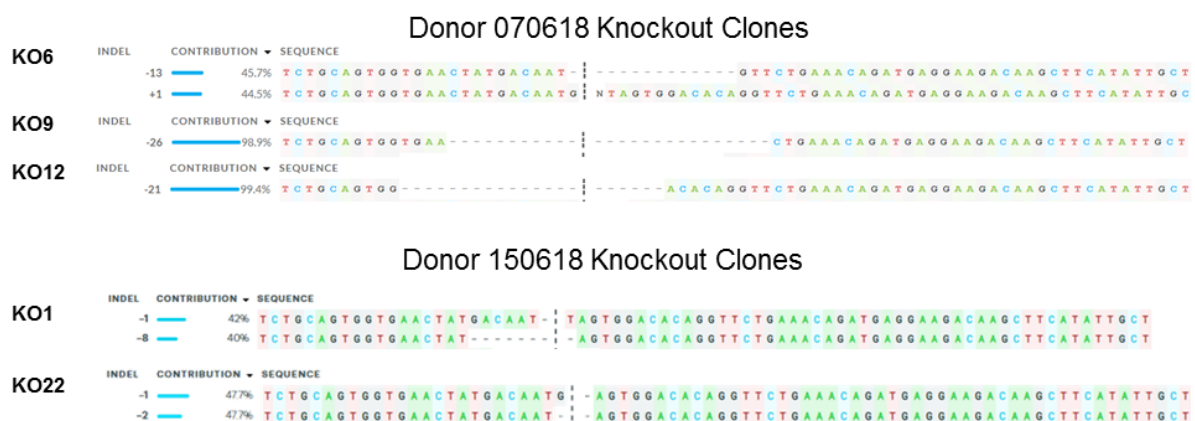
### [Genotype and expression profiles of ZEB2 deleted clones](#)

The variability in CRISPR/Cas9 editing, resulted not only in variable efficiency as described above, but also in differing types of editing (see Figure 4.3 above). This resulted in multiple versions of edited clones and also non-edited clones within the cell pool. In order to identify successfully edited cells and thus maximise the chance of observing phenotypic changes in ZEB2 deleted cells, I decided to isolate single cell clones. Single cell clones were isolated by

flow cytometry (as described in Section 2.3.3) and each of the clones was then expanded until there were sufficient cells for downstream applications. To facilitate this, the Th1 EM ZEB2 KO or WT clones were cultured in the presence of “feeder” cells (Mitomycin C inactivated PBMC) to encourage growth. Typically, the clonal populations were expanded for 1-2 months and those that survived were used for analysis of gene expression by qRT-PCR and global changes in gene expression by RNA-seq differential analysis. Analysis of editing efficiency (ICE) was carried out (section 2.3.6) on clones from both donors #070618 and #150618. In donor #070618, there were 3 clones with verified ZEB2 deletion (Figure 4.4). Clone KO6 was a heterozygous knockout with one allele having a 13bp deletion and the other allele having a base pair insertion. ZEB2 knockout clones: KO9 and KO12 both contained homozygous deletions (of 26bp and 21bp respectively) (Figure 4.4). In donor #150618, there were only two verified knockout clones: KO1 and KO22. Both alleles from these clones had heterozygous INDEL (Figure 4.4).

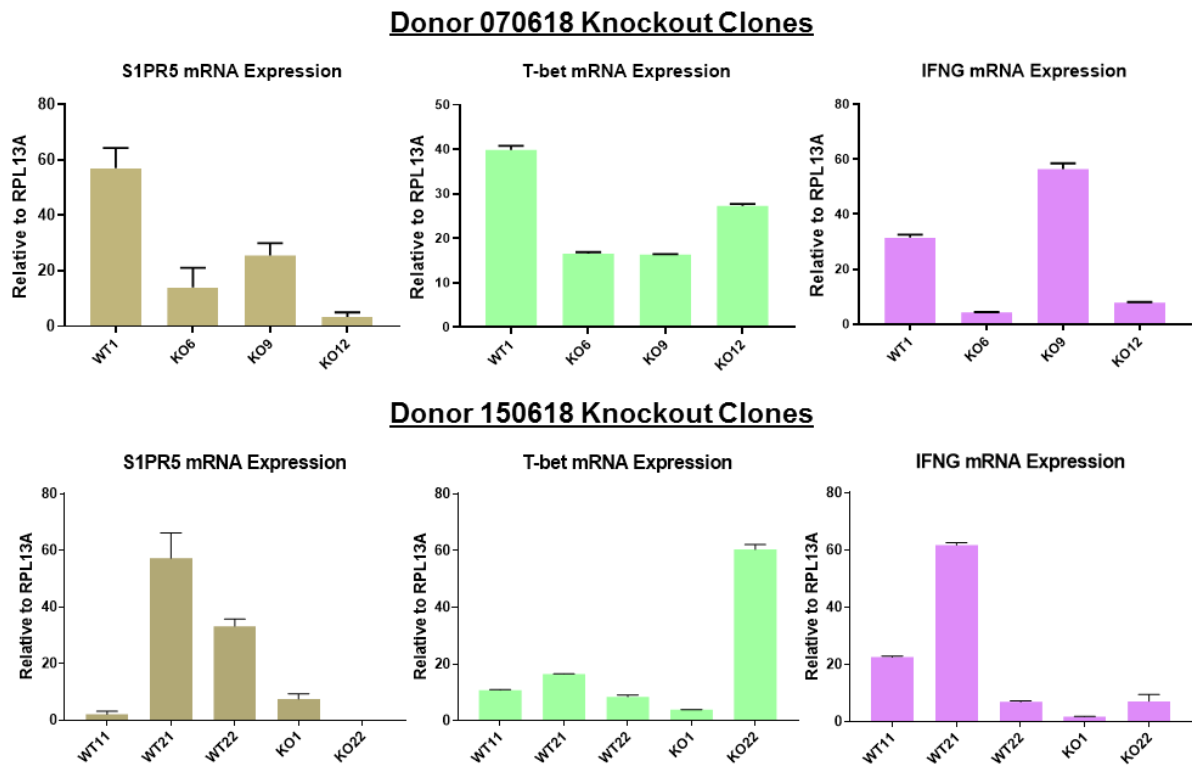
It has previously been shown that in NK and CD8 T cell where ZEB2 was ablated, expression of sphingosine 1-phosphate receptor 5 (S1PR5), a trafficking gene, was greatly reduced (Dominguez et al., 2015). Since I surmised that deletion of ZEB2 could potentially be seen only at the protein level, I decided to use expression of S1PR5 as marker for ZEB2 deletion in my Th1 EM clones. RNA was isolated from the ZEB2 KO and WT clones, and expression of S1PR5, T-bet and IFNG mRNA were analysed by qRT-PCR. The results between the two donors was highly variable. In donor #070618, expression of S1PR5 was lower in all of the KO clones compared with the WT clone. However, in donor #150618, there were three WT clones and each of these had very different expression of S1PR5. Two clones: WT21 and WT22 expressed high S1PR5, but clone WT 11 expressed low S1PR5 expression. This meant that although S1PR5 expression in the ZEB2 KO clones was low compared with the WT clones: WT21 and WT22, its expression in the two KO clones was higher than in WT11. WT11 may

have been an outlier but this result made interpretation difficult and the use of S1PR5 as a marker for ZEB2 knockout difficult. In addition, expression of Th1 master transcription factor T-bet and Th1 cytokine IFNG were inconsistent: in donor #070618, T-bet was highly expressed in the WT clone compared with the ZEB2 KO clones, but in donor #150618 ZEB2 KO1 clone had low T-bet expression compared with the 3 WT clones, but clone KO22 had much higher T-bet expression compared with the three WT clones. Expression of IFNG was likewise inconsistent, in donor #070618 where two of the 3 KO clones (KO6 and KO12) had lower IFNG expression compared with the WT clone, but clone KO9 had much higher expression of IFNG than the WT clone. In donor #150618 IFNG was low in both ZEB2 KO clones, but its expression was highly variable in the three WT clones (see Figure 4.4 below). Therefore, my results suggested that the differences and variability in gene expression of S1PR5, T-bet and IFNG may not have been necessarily owing to loss of ZEB2 but rather to clonal differences. I decided to exclude clone KO22 from downstream RNA-seq library processing because of its particularly high T-bet expression compared with all of the other clones of both KO and WT phenotype.



**Figure 4.4: Genotype of ZEB2 KO clones by ICE analysis**

Data from ICE analysis inferred sequences present in the edited KO clones (“Sequences”) and their relative representation in the edited KO pool (“Contribution”) are shown for donor #070618 (top) and donor #150618 (bottom). For each sequence, the number of nucleotides inserted (+) or deleted (-) was indicated under “INDEL”. The black vertical dotted line represents the cut site, and “+” symbol on the far left marks the wildtype.



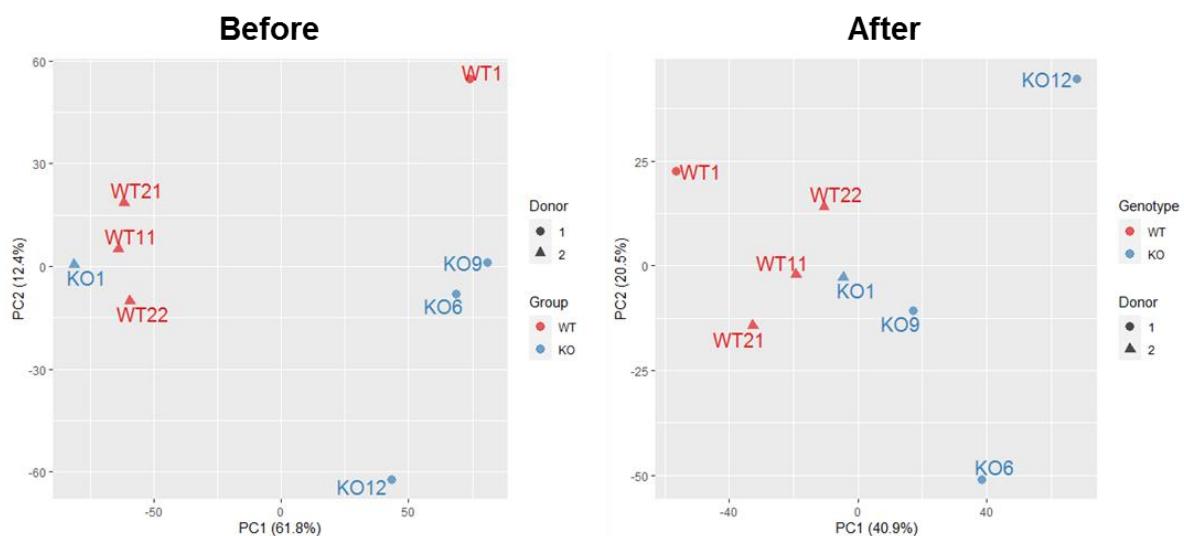
**Figure 4.5: Gene expression changes of ZEB2 KO Th1 EM clones by qRT-PCR**

SIPR5 (blue), T-bet (green) and IFNG (purple) expression in ZEB2 knockout clones. Th1 EM cells were nucleofected™ with single sgRNAs and then underwent clonal expansion. Donor 070618 (top) has one WT clone and three KO clone whereas donor 150818 (bottom) has three WT clones and one KO clone. Relative abundance of SIPR5, T-bet and IFNG mRNAs were normalised to reference gene RPL13A and plotted with mean and SEM from assay triplicates. No statistical tests were carried out for this experiment.

#### [Comparison of gene expression in Th1 EM ZEB2 knockout and wild-type clones by RNA-seq](#)

Global expression changes as a consequence of ZEB2 knockout in the Th1 EM clones were examined by RNA-sequencing (as described in Section 2.4). First, Principle Component Analysis (PCA) was carried out to examine low resolution differences in the RNA seq data in a low-dimensional subspace. The first thing I observed was that the samples were not clustered by their experimental groups, but rather by donor, as seen in Figure 4.6. RUVg normalisation was performed, using an empirical control with a k value of 1, in order to remove unwanted variation from the samples. Once normalisation of samples had been established, the WT clones clustered more closely together. However, KO clones were scattered around in the PCA space with KO1 and KO9 in close space with the WT clone experimental group (Figure 4.6). To

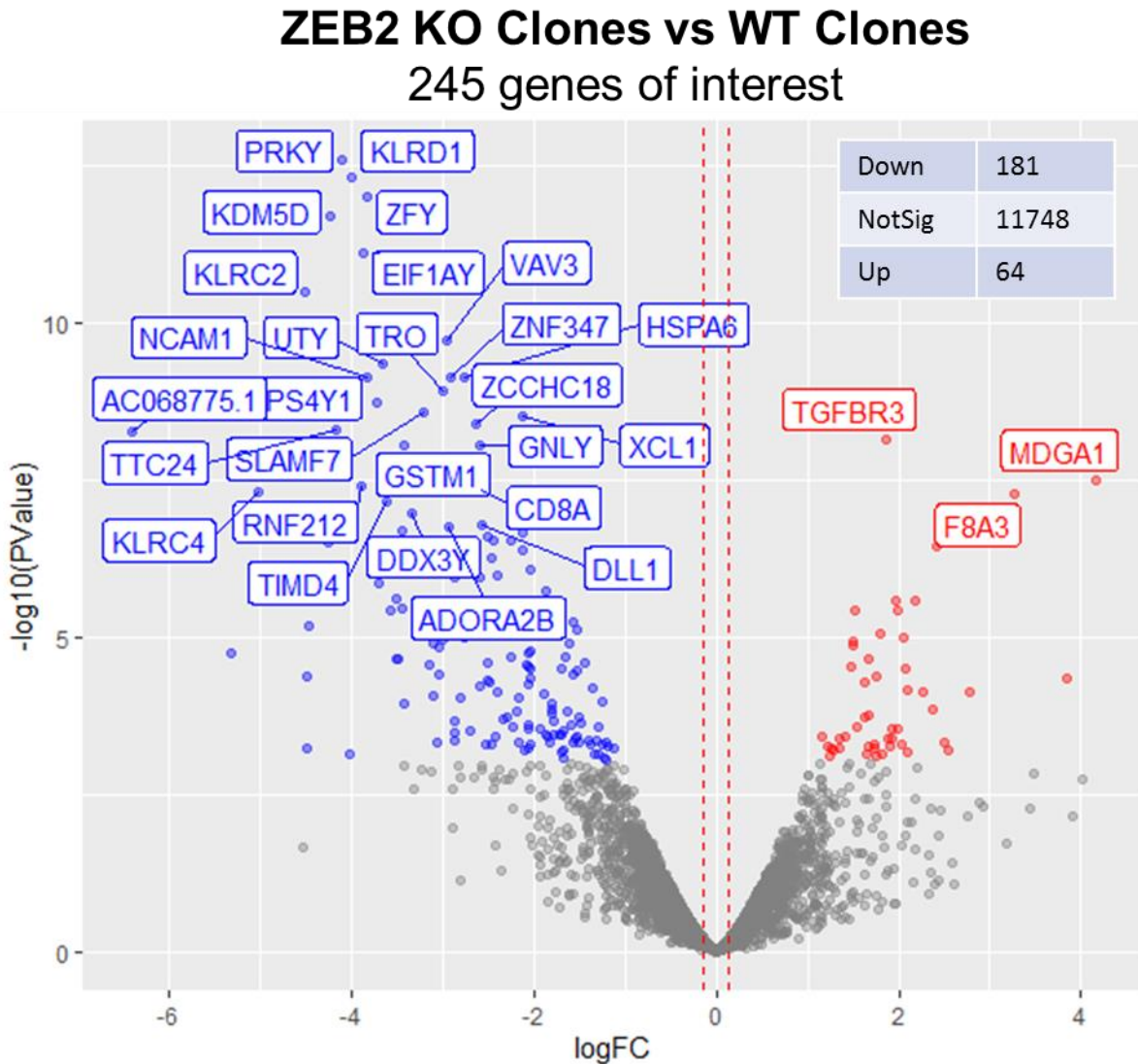
determine the ZEB2 regulated gene signatures, differential gene expression analysis was carried out (Figure 2.4.6). 245 genes were differentially expressed (DE) between ZEB2 KO and WT clones with 64 genes upregulated and 181 gene downregulated (Figure 4.7). Table 4.1 and Table 4.2 shows the top 35 upregulated and downregulated genes respectively. Despite ZEB2 being a transcriptional repressor, it was surprising that the loss of ZEB2 led to more genes being downregulated than upregulated. Clones were also grouped to their own experimental genotype based on the heatmap generated by unsupervised clustering using Euclidean methods (Figure 4.8). Together, these data indicate that ZEB2 KO in Th1 EM cell clones results in profound changes in gene networks potentially regulated by ZEB2.



**Figure 4.6: PCA of ZEB2 KO & WT Th1 EM clones before and after RUVg normalisation**

Principal component analysis of all samples showing genes before (left) and after (right) RUVg normalisation (k=1).





**Figure 4.7: Volcano plot comparing ZEB2 KO vs WT Th1 EM clones**

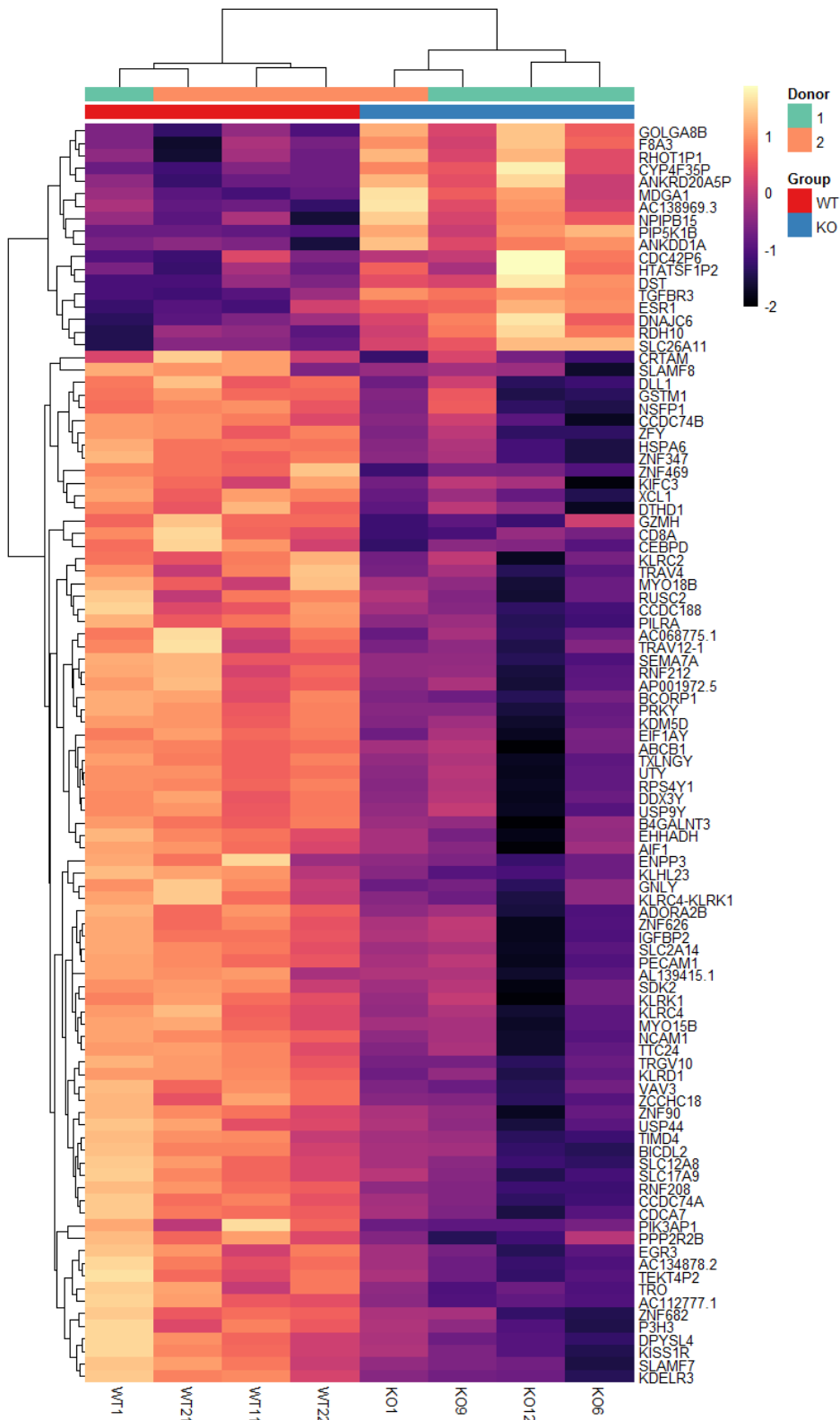
Volcano plot displaying differentially expressed genes between ZEB2 deleted and wildtype Control Th1 EM clones. The vertical axis (y-axis) corresponds to the mean expression value of  $\log_{10}(\text{PValue})$ , the horizontal axis (x-axis) displays the  $\log_2$  fold change value, the red vertical dotted lines indicate  $\log_2$  fold change of 1.1. The red dots represent the up-regulated expressed transcripts; the blue dots represent the transcripts whose expression is downregulated. The top 30 genes that were differentially expressed are labelled.

**Table 4.1: Top 35 upregulated genes in ZEB2 KO Th1 EM clones**

Gene Name	Gene Description	Log Fold-change	FDR adjusted P-value
TGFBR3	transforming growth factor beta receptor 3	1.88	3.36E-06
CYP4F35P	cytochrome P450 family 4 subfamily F member 35, pseudogene	2.58	1.06E-05
MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	4.19	1.39E-05
F8A3	coagulation factor VIII associated 3	3.31	2.39E-05
ESR1	estrogen receptor 1	2.43	8.04E-05
ANKRD20A5P	ankyrin repeat domain 20 family member A5, pseudogene	2.08	2.47E-04
SLC26A11	solute carrier family 26 member 11	1.98	4.26E-04
RDH10	retinol dehydrogenase 10	2.20	4.26E-04
AC138969.3	polycystic kidney disease 1 (autosomal dominant) (PKD1) pseudogene	2.80	5.00E-04
DNAJC6	DnaJ heat shock protein family (Hsp40) member C6	2.01	5.45E-04
PIP5K1B	phosphatidylinositol-4-phosphate 5-kinase type 1 beta	1.54	5.50E-04
HTATSF1P2	HIV-1 Tat specific factor 1 pseudogene 2	2.75	9.55E-04
DST	dystonin	1.83	1.16E-03
RHOT1P1	ras homolog family member T1 pseudogene 1	2.37	1.31E-03
NPIP15	nuclear pore complex interacting protein family member B15	2.06	1.36E-03
ANKDD1A	ankyrin repeat and death domain containing 1A	1.52	1.37E-03
CDC42P6	cell division cycle 42 pseudogene 6	3.63	1.54E-03
GOLGA8B	golgin A8 family member B	1.50	1.66E-03
CRYBG3	crystallin beta-gamma domain containing 3	1.68	2.16E-03
PLEC	plectin	1.49	2.76E-03
ADCY1	adenylate cyclase 1	2.13	3.05E-03
AKR1C3	aldo-keto reductase family 1 member C3	3.88	3.35E-03
TRAV23DV6	T cell receptor alpha variable 23/delta variable 6	6.81	3.56E-03
ASAH2	N-acylsphingosine amidohydrolase 2	1.76	3.81E-03
NQO1	NAD(P)H quinone dehydrogenase 1	1.64	4.13E-03
PPARG	peroxisome proliferator activated receptor gamma	2.28	5.07E-03
STEAP1B	STEAP family member 1B	2.12	5.54E-03
FOXP3	forkhead box P3	2.79	5.54E-03
AC243919.1	ribosomal protein L9 pseudogene 9	1.88	6.72E-03
TLR6	toll like receptor 6	2.39	1.01E-02
SLC35F3	solute carrier family 35 member F3	1.68	1.15E-02
DOCK3	dedicator of cytokinesis 3	1.62	1.43E-02
CYP7B1	cytochrome P450 family 7 subfamily B member 1	1.97	1.49E-02
AP005212.2	sorting nexin 18 (SNX18) pseudogene	2.34	1.58E-02
CD28	CD28 molecule	1.54	1.82E-02

**Table 4.2: Top 35 downregulated genes in ZEB2 KO Th1 EM clones**

Gene Name	Gene Description	Log Fold-change	FDR adjusted P-value
TRGV10	T cell receptor gamma variable 10 (non-functional)	-6.17	2.10E-14
KLRD1	killer cell lectin like receptor D1	-4.09	1.42E-09
PRKY	protein kinase Y-linked (pseudogene)	-4.16	1.45E-09
ZFY	zinc finger protein Y-linked	-3.89	2.62E-09
KDM5D	lysine demethylase 5D	-4.29	7.31E-09
EIF1AY	eukaryotic translation initiation factor 1A Y-linked	-3.93	1.94E-08
KLRC2	killer cell lectin like receptor C2	-4.57	1.94E-08
BCORP1	BCL6 corepressor pseudogene 1	-3.95	2.94E-08
VAV3	vav guanine nucleotide exchange factor 3	-2.97	1.78E-07
ZNF347	zinc finger protein 347	-2.94	7.00E-07
UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	-3.72	7.00E-07
HSPA6	heat shock protein family A (Hsp70) member 6	-2.77	7.00E-07
TXLNGY	taxilin gamma pseudogene, Y-linked	-3.92	7.00E-07
NCAM1	neural cell adhesion molecule 1	-3.83	7.00E-07
TRO	trophinin	-3.03	7.29E-07
SLAMF7	SLAM family member 7	-3.25	1.46E-06
XCL1	X-C motif chemokine ligand 1	-2.13	1.57E-06
ZCCHC18	zinc finger CCHC-type containing 18	-2.65	1.86E-06
RPS4Y1	ribosomal protein S4 Y-linked 1	-3.74	1.86E-06
TTC24	tetratricopeptide repeat domain 24	-4.25	2.18E-06
AC068775.1	novel protein	-6.51	2.44E-06
AC112777.1	ubiquitin-like with PHD and ring finger domains 1 (UHRF1) pseudogene	-2.77	3.57E-06
GNLY	granulysin	-2.61	4.34E-06
TRAV4	T cell receptor alpha variable 4	-3.86	4.34E-06
GSTM1	glutathione S-transferase mu 1	-3.47	5.64E-06
TRAV12-1	T cell receptor alpha variable 12-1	-5.68	6.90E-06
CD8A	CD8a molecule	-2.64	8.81E-06
RNF212	ring finger protein 212	-3.97	1.30E-05
KLRC4	killer cell lectin like receptor C4	-5.14	1.49E-05
NSFP1	N-ethylmaleimide-sensitive factor pseudogene 1	-2.84	2.08E-05
TIMD4	T cell immunoglobulin and mucin domain containing 4	-3.62	2.39E-05
ADORA2B	adenosine A2b receptor	-3.00	3.15E-05
AC134878.2	MAFF interacting protein (pseudogene)	-4.41	3.96E-05
DDX3Y	DEAD-box helicase 3 Y-linked	-3.37	4.79E-05
DLL1	delta like canonical Notch ligand 1	-2.60	4.80E-05



**Figure 4.8: Heatmap of top 100 DE genes between ZEB2 KO and WT Th1 EM clones**

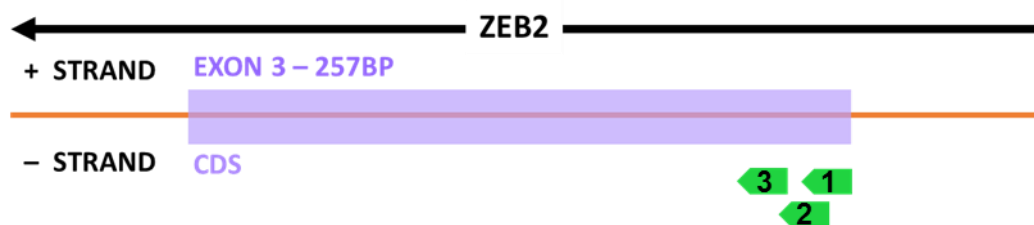
Hierarchical clustering heatmap analysis of the top 100 most differentially expressed genes between ZEB2-KO and WT Th1 EM clones. Heatmaps are based on log-transformed expression values, are z-scaled by rows and were plotted using the R package pheatmap. All genes and samples were clustered with Euclidean distance with  $n = 4$  per group.

Overall, the results from using single the sgRNA approach for deleting ZEB2 has indicated that not only was transfection efficiency highly variable, but also that the efficiency of successful gene editing was highly variable. One explanation for this could be that the use of a single CRISPR guide results in variable editing efficiencies in a pool of cells, suggesting a need to further optimise the CRISPR/Cas9 system.

#### 4.4.3 A Triple sgRNAs CRISPR/Cas9 Deletion of ZEB2

I found that the knockout efficiency was highly variable using the single sgRNA system in CRISPR/Cas9 deletion of ZEB2, also reported by other studies using the single sgRNAs approach (Schumann et al., 2015) suggesting that the single sgRNA format is not always efficient. However, a more recent studies by Seki and Rutz (2018) tested a triple sgRNAs approach and observed an increased loss of protein expression.

Therefore, to increase the efficacy of CRISPR-mediated ZEB2 knockout in Th1 EM cells, I have adopted the triple sgRNAs approach. For these experiments, two additional sgRNAs (Table 2.4) were designed and used together with the initial single sgRNA, targeting regions of ZEB2 exon 3 in close proximity to each other (Figure 4.9). This approach to targeting can cause multiple concurrent double-stranded breaks in the genomic DNA so that a large fragment is removed. Because this type of editing is so disruptive, there is a high likelihood that the targeted gene will be rendered inoperative.



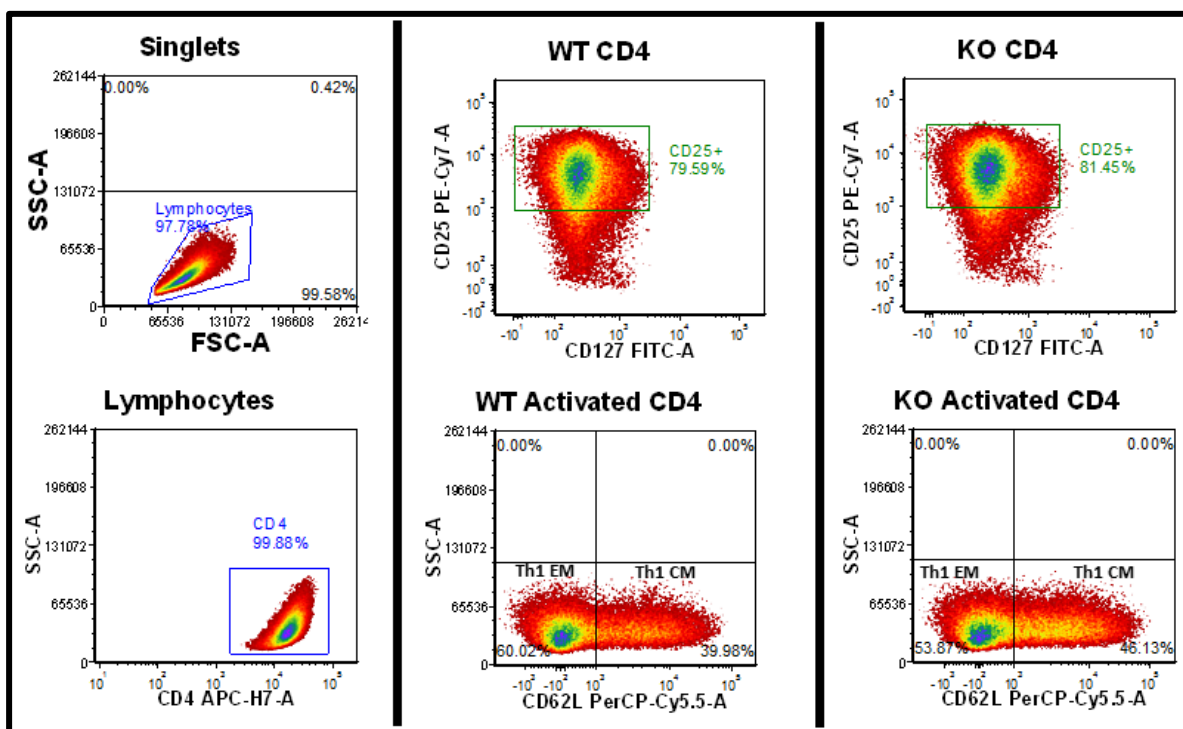
**Figure 4.9: Schematic of 3 sgRNAs targeting exon 3 of ZEB2 gene**

#### [Analysis of ZEB2 knockout in pooled Th1 EM cells using 3 sgRNAs strategy](#)

As well as using a three sgRNA CRISPR strategy instead of a single sgRNA strategy, I also decided to change the cell isolation and culture conditions. This was carried out in an attempt to increase efficiency and viability of the cells post Nucleofection™ and also to prevent artefacts associated with the long-term culture required with the single cell cloning strategy. I decided to first isolate Th1 cells rather than Th1 EM since this results in higher cell numbers (>2x10<sup>6</sup> cells). It is worth noting that I carried out two independent KO nucleofection™

experiments each for donor #080119 and donor #100119 since I obtained very high cell numbers with these donors. After Nucleofection™, the CRISPR edited and non-edited cells were then separated by flow cytometry and gated on CD25+ (activated) cells, activated cells being more likely to endocytose the CRISPR/Cas9 RNPs complex. The CD25+ cells were then segregated into the EM population by FACS with stringent gating as described in Section 2.1.6, but this time I also retained the CM population for comparative study (based on their CD62L status by flow cytometry) (Figure 4.10). After isolation of genomic DNA from the ZEB2 KO and WT cells, the first thing I observed was that the amplicons from the PCR amplification of the ZEB2 targeted region, were much more diffuse in the ZEB2 KO cells using the three guide sgRNAs compared with using the single sgRNA system (Figure 4.11 compared to Figure 4.2). In addition, ICE analysis of the Th1 EM ZEB2 KO showed increased INDEL and KO efficiency overall (

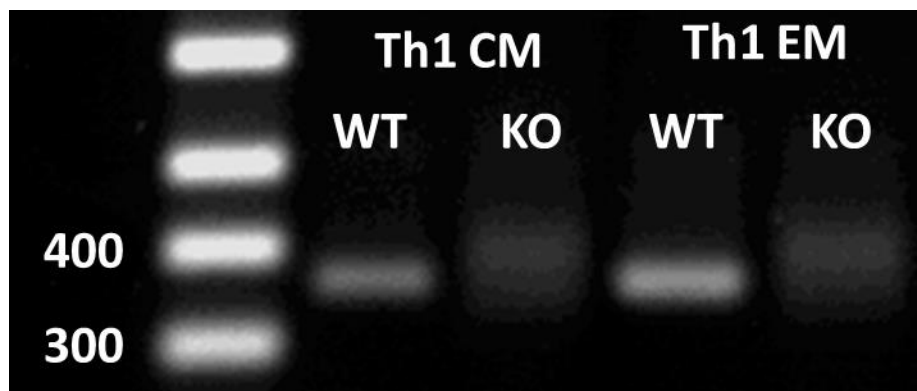
Table 4.3) compared with Figure 4.3 where a single sgRNA was used. In all of the donors, the efficiency of ZEB2 INDELS in the Th1 EM cells was higher than 75% and importantly, the knockout efficiency was consistently higher (at least 50%). Overall, these results indicated that this approach of using triple sgRNAs in Th1 pools achieved high editing efficiency for downstream application, without the need to carry out time consuming clonal expansion. The entire experimental timeline was effectively shortened from the initial 2-4 months (single sgRNAs approach) to 1 week (triple sgRNAs approach) for each donor.



**Figure 4.10: Segregation of post-nucleofected Th1 into CM and EM**

FACS plot showing Th1 CM (CD62L+) and Th1 EM (CD62L-) gated and then isolated from both WT and KO Th1 cell pools. Purified Th1 cells were first activated for 3 days prior to delivery of CRISPR/Cas9 via Nucleofection™. Post-Nucleofected cells were rested for 2 days and stained for CD4 (cell population identification), CD25 (activation marker) and CD62L (memory status). Physical gating is shown in the left panel, gating for Th1 CM and EM cells from WT and KO are shown in middle and right panel respectively. Representative data from 1 donor.





**Figure 4.11: Gel image of genomic PCR post CRISPR/Cas9 using three sgRNAs**

Post genomic PCR of WT and ZEB2-KO Th1 CM and EM were electrophoresed using 1% agarose and 80V. Representative data from 1 donor.

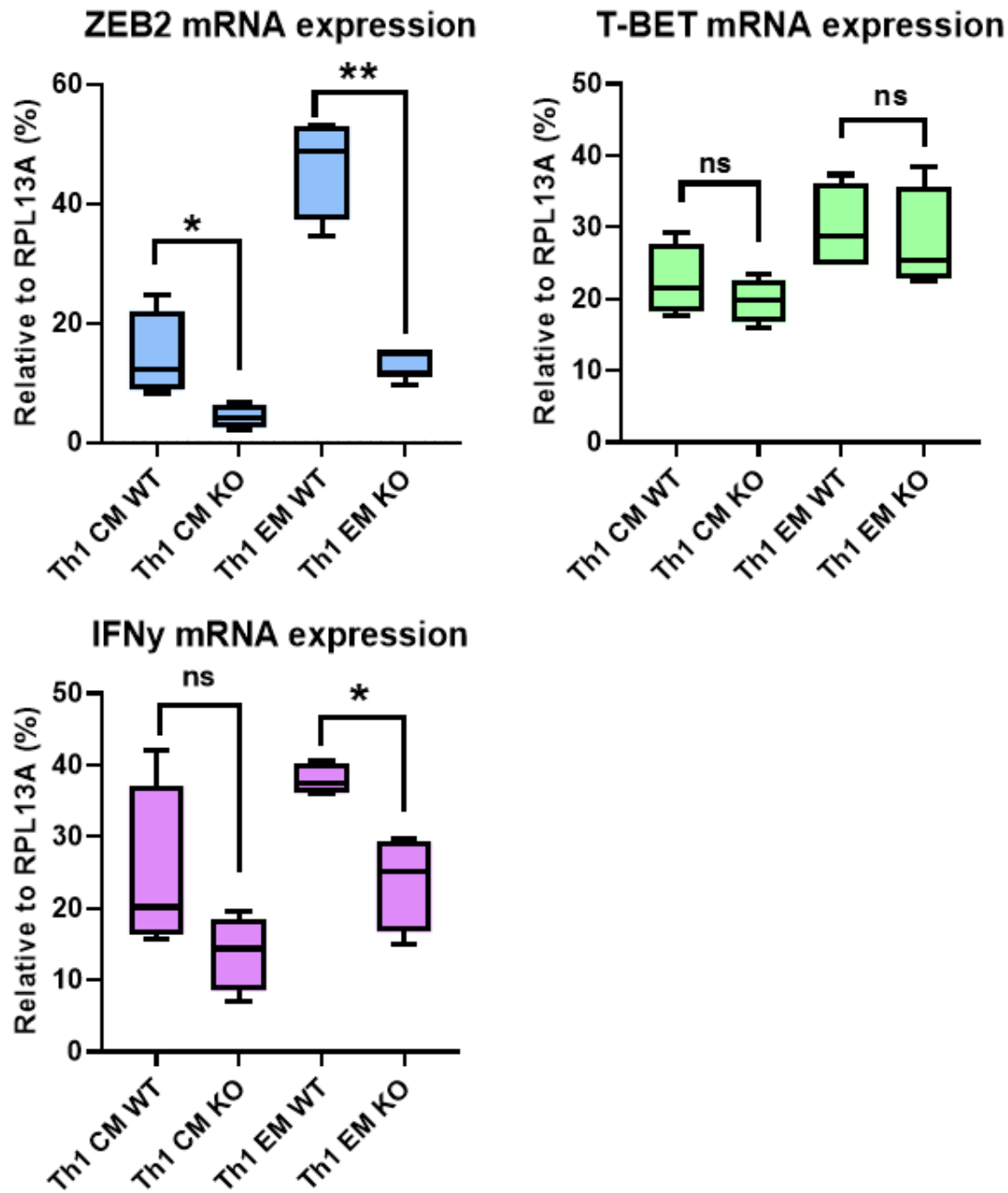
**Table 4.3: Summary of ICE analysis of triple sgRNAs ZEB2 KO Th1 EM**

Label	ICE (%)	KO-Score (%)	R Squared	Mean Discord Before	Mean Discord After
080119 KO EM1	74	64	0.74	0.275	0.748
080119 KO EM2	76	54	0.93	0.177	0.596
160119 KO EM	85	59	0.91	0.195	0.675
100119 KO EM2	78	55	0.93	0.194	0.606
100119 KO EM1	74	54	0.93	0.23	0.586
180119 KO EM	83	61	0.93	0.143	0.65

#### ZEB2 is required for the expression of Th1 effector cytokine, IFN $\gamma$

Before proceeding with RNA-seq analysis of the gene networks impacted by knockout of ZEB2, I wished to determine if this new approach for ZEB2 knockout, using the 3 sgRNA CRISPR strategy and Th1 EM cell pools rather than clones, would enable alterations in gene expression to be detected by RT PCR in the ZEB2 KO cells compared with WT cells. The results of this were very interesting, expression of ZEB2 mRNA was clearly reduced in ZEB2 KO cells compared with WT cells, indicating the robustness of the deletion strategy (Figure 4.12). Since Th1 CM populations were also retained during the sort, I examined if ZEB2 knockout also has an effect on Th1 CM cells. Here, loss of ZEB2 was observed both in EM cells and also in CM cells, although expression of T-bet and IFN $\gamma$  were statistically significantly altered by the knockout (Figure 4.12). Interestingly, in the ZEB2 deleted Th1 EM cells, expression of the Th1 defining transcription factor, T-bet was not altered, indicating that deletion of ZEB2 does not have an impact on expression of T-bet and therefore suggesting that T-bet expression is not dependent on ZEB2, supporting the findings of Omilusik et al. (2015), Dominguez et al. (2015) and van Helden et al. (2015). IFN $\gamma$  mRNA expression, however, was significantly reduced in the Th1 EM ZEB2 KO cells compared with WT cells. This important finding suggests that ZEB2 may have a novel role in regulating IFN $\gamma$  in Th1 EM cells but to confirm this in the Th1 CM subset the sample size would need to be increased to address the

variance in the results (Figure 4.12). IFN $\gamma$  is absolutely required for Th1 effector function and so loss of this effector cytokine would lead to reduced Th1 potency. This result shows that ZEB2 is required for Th1 EM effector function and has significant implications for the efficacy of pathogen clearance by Th1 cells.



**Figure 4.12: Gene expression changes in ZEB2 KO Th1 CM and EM cells by qRT-PCR**

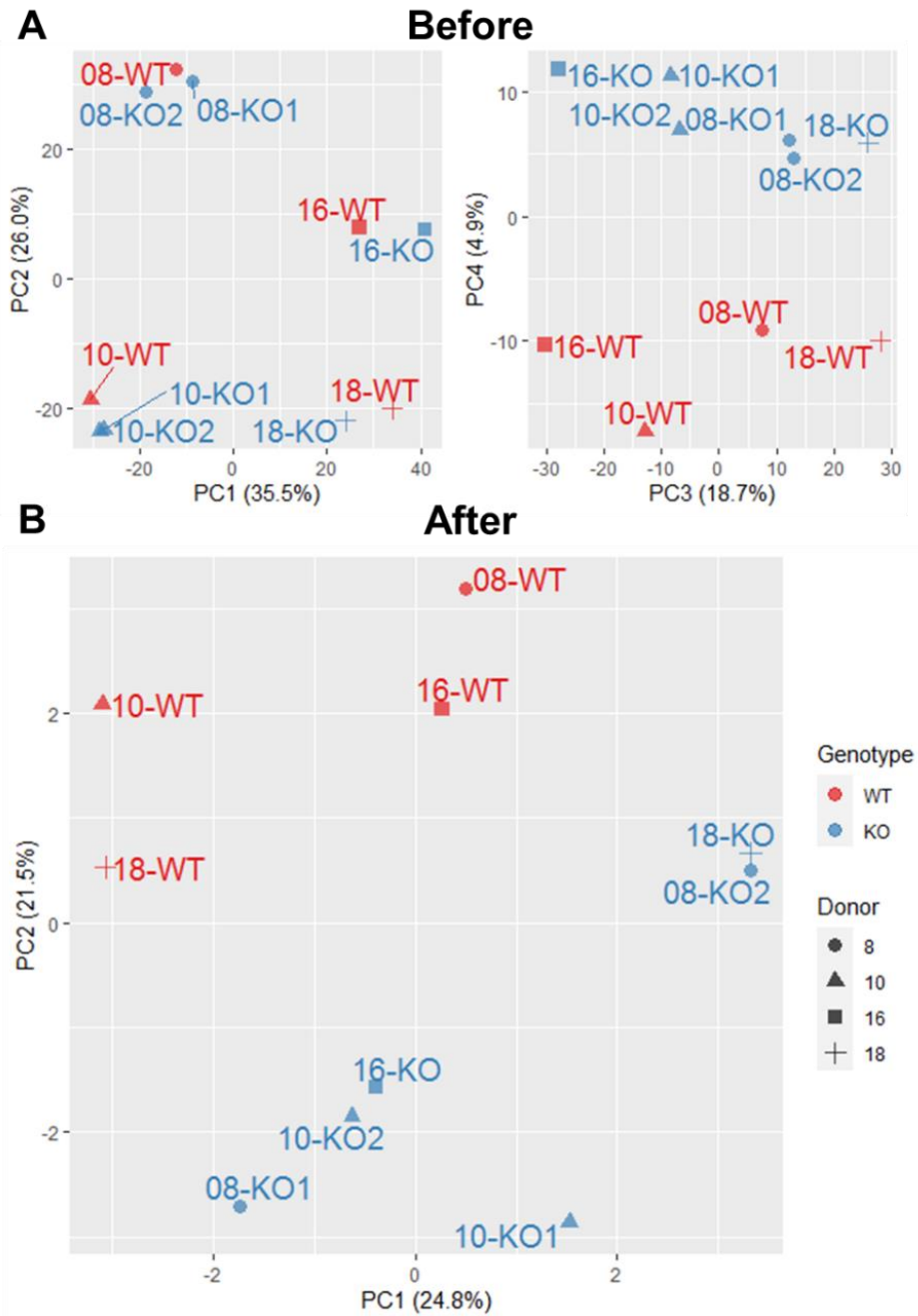
Th1 cells were Nucleofected™ with triple sgRNAs to delete ZEB2 and cells were gated on CD25+ and separated them into CM (CD62L+) and EM (CD62L-) cells. Expression of ZEB2 (blue), T-bet (green) and IFN $\gamma$  (purple) mRNAs was examined in ZEB2 KO and WT of Th1 CM and Th1 EM cells. Relative abundance of ZEB2, T-bet, FOXP3, and IFN $\gamma$  were normalised to reference gene RPL13A and plotted with box and whisker, statistics were carried out with paired t test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ,  $n = 3$  independent donors.

### Th1 lineage fidelity is maintained by ZEB2

Since the previous experiments have shown that ZEB2 is highly likely to be required for optimal function of Th1, and as ZEB2 is a transcription factor, it may directly or indirectly influence the transcription of multiple genes which may shape Th1 function and fidelity. To gain insight into the mechanism of action and potential downstream targets of ZEB2 in Th1 EM, I performed RNA-seq on WT and ZEB2-KO Th1 EM pools to look for changes in global gene expression. First, I carried out Principal Component Analysis (PCA) to see how the populations of cells clustered, which gives an idea of how similar the populations are to each other. The principal component axis (PC) assumes the directions from largest to smallest variance. For example, the PC1 axis is the first principal direction along which the samples show the largest variation, whereas the PC2 axis is the second most important direction, and the same applies further down the PC3 and PC4 axis with less contribution to the variation.

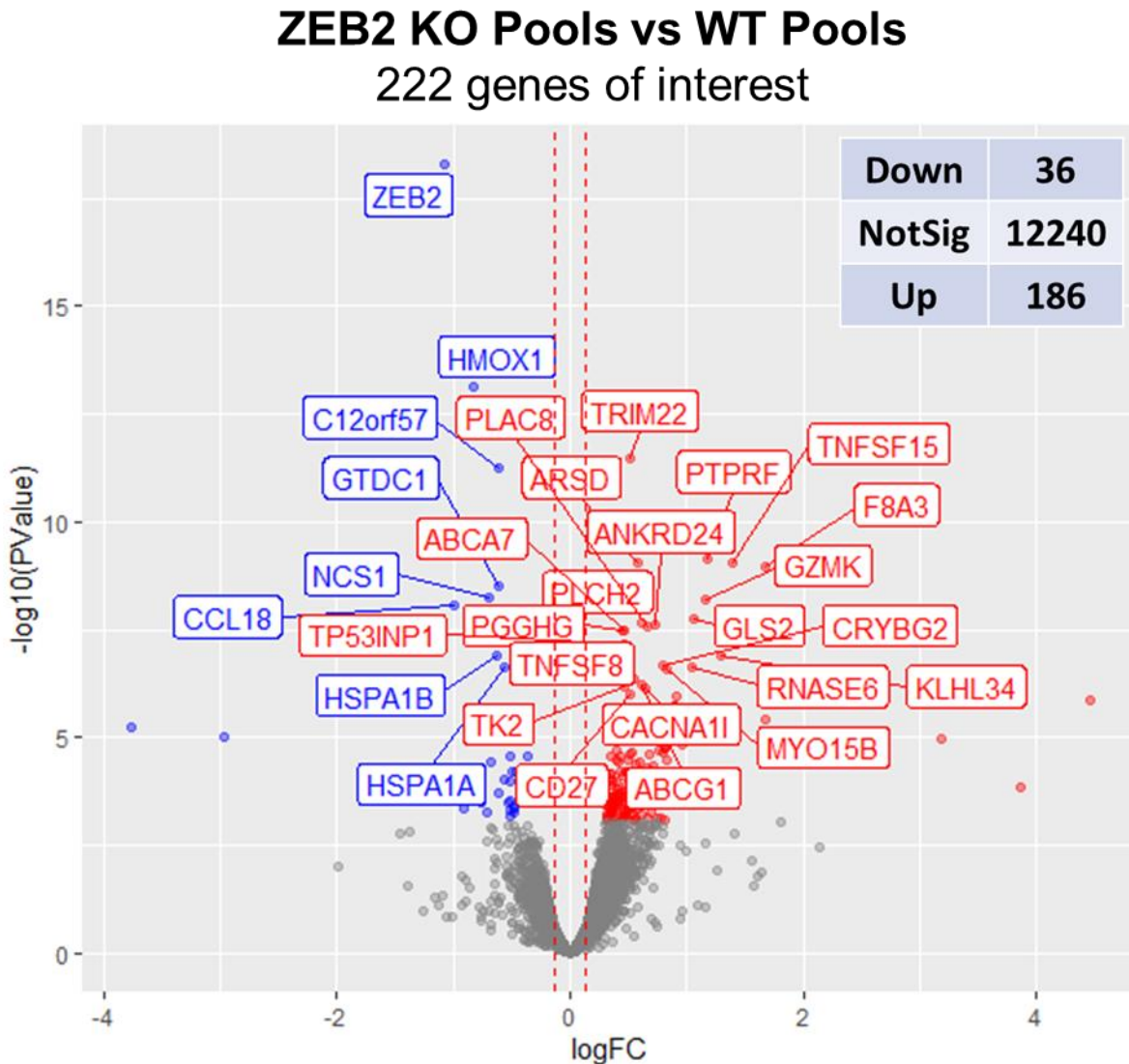
As seen in Figure 4.13A, samples were clustered based on donors in the PC1 and PC2 axes, suggesting that the strongest variance and difference between the 12 datasets is the donor, not the experimental group. It was only at the PC3 and PC4 levels, sample clustering by experimental group was observed. This type of clustering relationship suggested that there was a requirement for removal of donor variation prior to Differential Expression (DE) analysis. RUVg normalisation was therefore carried out to remove unwanted variation by using a parameter of  $k=3$  on a set of negative control genes (least significant DE genes) which are assumed to have constant expression across all samples. After normalisation, samples now cluster by experimental WT and KO groupings (Figure 4.13B). Interestingly, the effect of ZEB2 deletion in both Th1 EM cells pool (Figure 4.13A) or clones (Figure 4.6) were not a major contributing factor to the overall phenotypic changes compared with donor differences, therefore RUVg normalisation was required in analysis of both datasets to normalise the donor variation.

An RNA sequencing (RNA-seq) differential expression analysis pipeline was carried out (described in the Section 2.4.6) using a False Discovery Rate (FDR) of  $< 0.05$  and a cut-off of 1.1 log<sub>2</sub>-fold change, generating a list of significantly differentially expressed (DE) genes between the knockout (KO) and wild type (WT) samples. The less stringent log<sub>2</sub>-fold change of 1.1 was used in order to capture the low differentially expressed genes signals owing to the signal being diluted by the presence of WT cells within the KO pools (cells were not 100% KO in the KO pool). In total, there were 222 significantly differentially expressed (DE) genes between the two groups (Figure 4.14). As ZEB2 is predominantly a transcriptional repressor; it is reassuring to observe that more genes were upregulated (186 genes) than downregulated (36 genes). Table 4.4 and Table 4.5 show the top 35 upregulated and downregulated genes respectively. To show consistency of expression patterns of the top Differentially Expressed genes, we used an unsupervised clustering heatmap of the top 100 genes from all donors (Figure 4.15). The major Th1 effector cytokine, IFN $\gamma$ , was also significantly reduced in the KO RNA-seq from all donors, consistent with our qPCR data (Table 4.5).



**Figure 4.13: PCA of ZEB2 KO & WT Th1 EM cell pools before and after RUVg normalisation**

Principal component analysis of all samples run on all expressed genes. (A) PC1-PC2 (left) showed samples clustered by donor and PC3-PC4 (right) showed samples clustered by experimental group. (B) PC1-PC2 showed samples clustered by donor after RUVg normalisation of  $k=3$ .



**Figure 4.14: Volcano plot comparing ZEB2 KO vs WT Th1 EM pools**

Volcano plot displaying differentially expressed genes between ZEB2 deleted and wildtype Control of Th1 EM pools. The vertical axis (y-axis) corresponds to the mean expression value of  $-\log_{10}(\text{PValue})$ , the horizontal axis (x-axis) displays the  $\log_2$  fold change value, the red vertical dotted lines indicate  $\log_2$  fold change of 1.1. The red dots represent the up-regulated expressed transcripts; the blue dots represent the transcripts whose expression is downregulated. The top 30 genes that were differentially expressed are labelled.

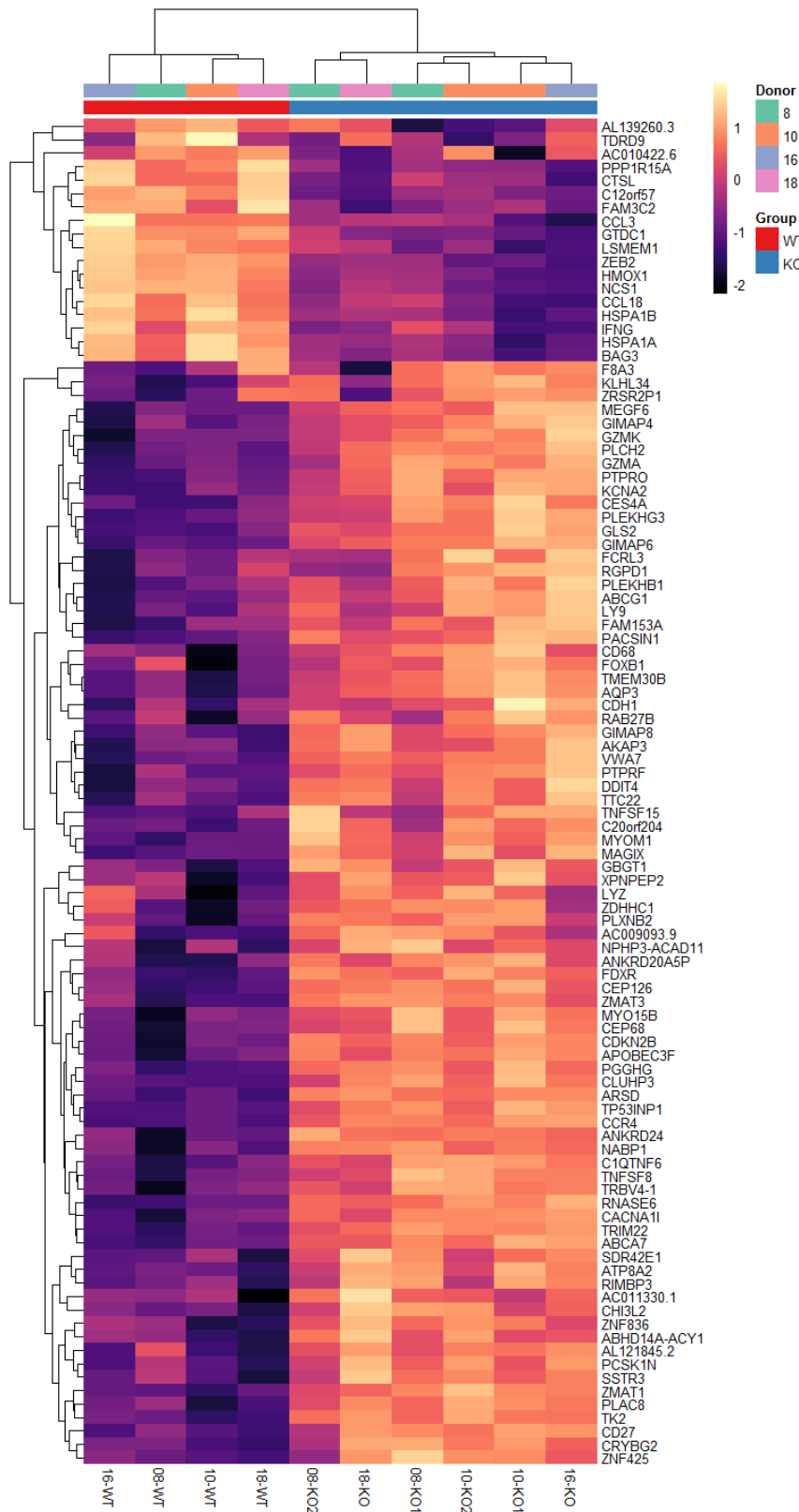


**Table 4.4: Top 35 upregulated genes in ZEB2 KO Th1 EM pools**

Gene Name	Gene Description	Log Fold-change	FDR adjusted P-value
TRIM22	tripartite motif containing 22	0.514896	1.36E-08
PTPRF	protein tyrosine phosphatase receptor type F	1.180568	1.63E-06
ARSD	arylsulfatase D	0.586349	1.63E-06
TNFSF15	TNF superfamily member 15	1.391549	1.63E-06
F8A3	coagulation factor VIII associated 3	1.681375	1.64E-06
GZMK	granzyme K	1.153786	7.35E-06
GLS2	glutaminase 2	1.054424	1.70E-05
PLCH2	phospholipase C eta 2	0.619423	1.94E-05
ANKRD24	ankyrin repeat domain 24	0.729956	2.10E-05
PLAC8	placenta associated 8	0.666133	2.10E-05
ABCA7	ATP binding cassette subfamily A member 7	0.457707	2.26E-05
PGGHG	protein-glucosylgalactosylhydroxylysine glucosidase	0.453396	2.26E-05
TP53INP1	tumor protein p53 inducible nuclear protein 1	0.468863	4.46E-05
KLHL34	kelch like family member 34	1.28974	7.80E-05
CRYBG2	crystallin beta-gamma domain containing 2	0.796606	0.000119
RNASE6	ribonuclease A family member k6	1.03825	0.000119
MYO15B	myosin XVB	0.826793	0.000129
TNFSF8	TNF superfamily member 8	0.550316	0.000219
ABCG1	ATP binding cassette subfamily G member 1	0.611515	0.000283
TK2	thymidine kinase 2	0.459949	0.0003
CACNA1I	calcium voltage-gated channel subunit alpha1 I	0.647652	0.000326
CD27	CD27 molecule	0.522125	0.000406
ATP8A2	ATPase phospholipid transporting 8A2	0.91731	0.000471
ZRSR2P1	ZRSR2 pseudogene 1	4.466846	0.000538
FAM153A	family with sequence similarity 153 member A	0.890096	0.000677
LYZ	lysozyme	0.899434	0.000703
AKAP3	A-kinase anchoring protein 3	0.672526	0.000703
MYOM1	myomesin 1	0.677391	0.000742
C20orf204	chromosome 20 open reading frame 204	0.532484	0.000909
PTPRO	protein tyrosine phosphatase receptor type O	1.037709	0.000909
ZNF836	zinc finger protein 836	0.61347	0.001019
PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	0.629405	0.001097
CCR4	C-C motif chemokine receptor 4	0.345587	0.001114
VWA7	von Willebrand factor A domain containing 7	0.650253	0.001174
AC011330.1	histidine acid phosphatase domain containing 2A (HISPPD2A) pseudogene	1.673019	0.001174

**Table 4.5: Top 35 downregulated genes in ZEB2 KO Th1 EM pools**

Gene Name	Gene Description	Log Fold-change	FDR adjusted P-value
ZEB2	zinc finger E-box binding homeobox 2	-1.06974	6.37E-15
HMOX1	heme oxygenase 1	-0.82657	4.69E-10
C12orf57	chromosome 12 open reading frame 57	-0.61429	1.78E-08
GTDC1	glycosyltransferase like domain containing 1	-0.62005	4.16E-06
NCS1	neuronal calcium sensor 1	-0.69221	6.94E-06
CCL18	C-C motif chemokine ligand 18	-0.98659	8.57E-06
HSPA1B	heat shock protein family A (Hsp70) member 1B	-0.62683	7.77E-05
HSPA1A	heat shock protein family A (Hsp70) member 1A	-0.55526	0.000119
AC010422.6	novel transcript	-3.76577	0.001549
LSMEM1	leucine rich single-pass membrane protein 1	-0.8283	0.001798
AL139260.3	novel protein	-2.97487	0.002073
PPP1R15A	protein phosphatase 1 regulatory subunit 15A	-0.36866	0.004443
CTSL	cathepsin L	-0.50954	0.004448
CCL3	C-C motif chemokine ligand 3	-0.67944	0.005441
BAG3	BAG cochaperone 3	-0.36872	0.007266
FAM3C2	family with sequence similarity 3 member C2 (pseudogene)	-0.45934	0.008522
IFNG	interferon gamma	-0.50076	0.008739
TDRD9	tudor domain containing 9	-0.88819	0.010417
DNAJB1	DnaJ heat shock protein family (Hsp40) member B1	-0.43761	0.010723
AC092117.2	tec	-0.79928	0.011227
LCN10	lipocalin 10	-0.56598	0.011302
HSPA6	heat shock protein family A (Hsp70) member 6	-0.51807	0.011852
HSPE1	heat shock protein family E (Hsp10) member 1	-0.32177	0.015576
AC073264.3	novel TRPM8 channel-associated factor pseudogene	-1.01849	0.01818
SHC4	SHC adaptor protein 4	-0.61687	0.018776
CTTN	cortactin	-0.50897	0.023364
ZFAND2A	zinc finger AN1-type containing 2A	-0.378	0.023452
SIM2	SIM bHLH transcription factor 2	-0.75544	0.024929
NCR3LG1	natural killer cell cytotoxicity receptor 3 ligand 1	-0.53142	0.025722
HID1	HID1 domain containing	-0.48835	0.029452
NPIP6	nuclear pore complex interacting protein family member B6	-0.90832	0.029955
CBS	cystathionine beta-synthase	-0.47652	0.031861
SIPA1L2	signal induced proliferation associated 1 like 2	-0.52044	0.034523
PROK2	prokineticin 2	-0.48423	0.034823
BMP4	bone morphogenetic protein 4	-0.7154	0.034823



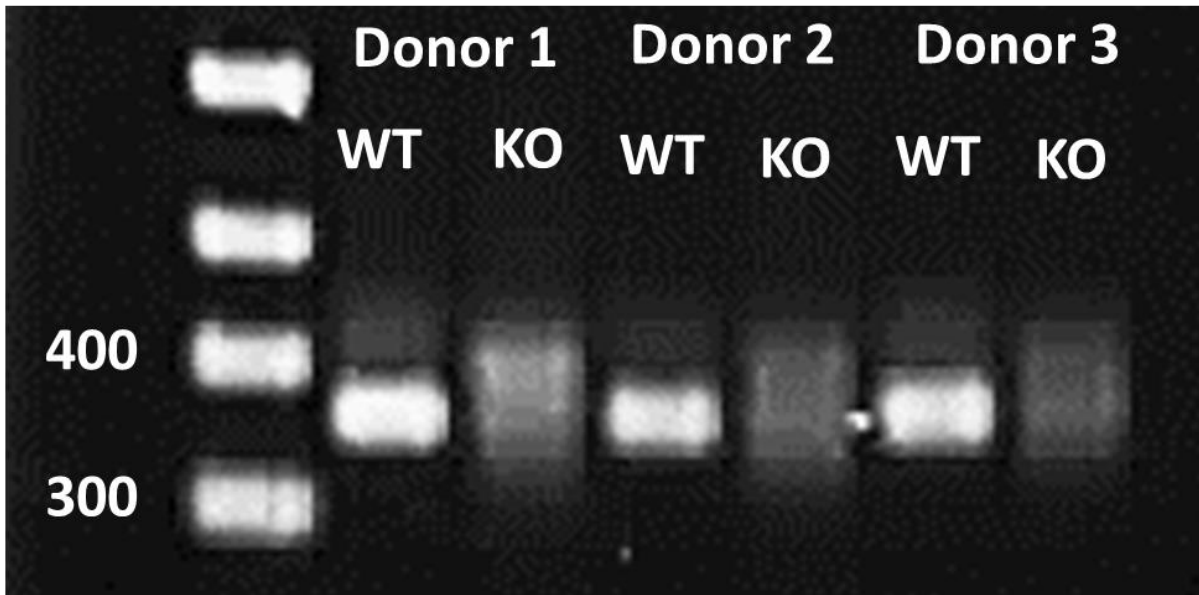
**Figure 4.15: Heatmap of top 100 DE genes between ZEB2 KO and WT Th1 pools**

Hierarchical clustering heatmap analysis of the top 100 most differentially expressed genes between ZEB2-KO and WT Th1 EM pools. Heatmaps are based on log-transformed expression values, are z-scaled by rows and were plotted using the R package pheatmap. All genes and samples were clustered with Euclidean distance.

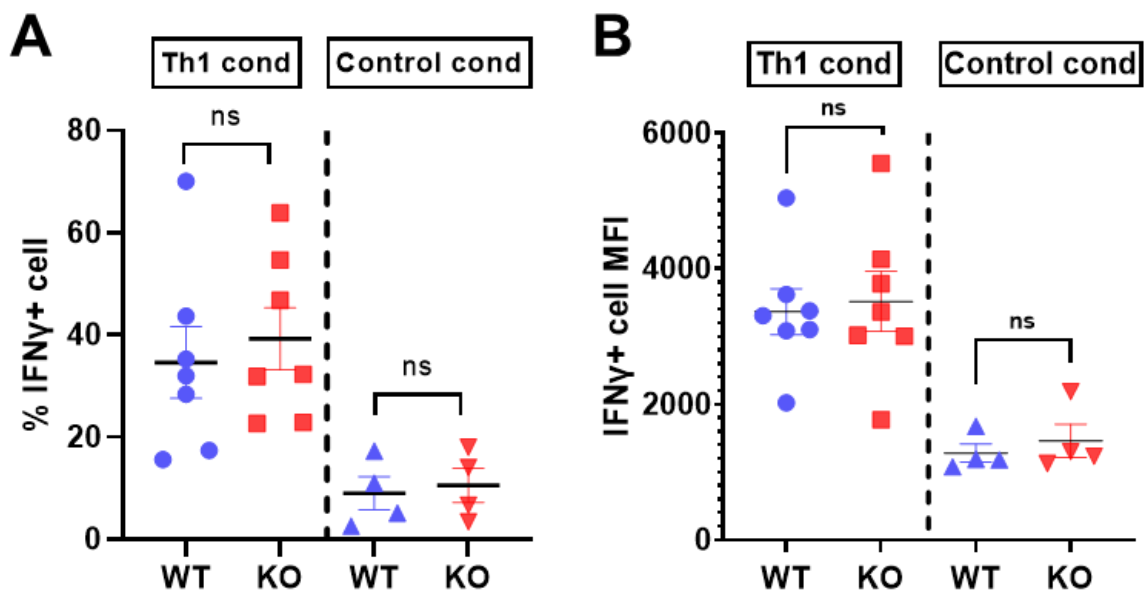
#### 4.4.4 ZEB2 is not required for formation of Th1

In the previous experiments, I have shown that ZEB2 is necessary for the expression of the Th1 cytokine IFN $\gamma$ , and therefore essential for the function of Th1 cells. I now wished to find out whether ZEB2 is also necessary for the formation of the Th1 cells. For these experiments I ablated ZEB2 in the Naïve Tconv using the triple sgRNAs, confirming the likelihood of successful deletion of ZEB2 by agarose gel electrophoresis of PCR amplicons as before (Figure 4.16). The ZEB2 KO and WT naïve CD4<sup>+</sup> T cells were cultured in Th1 polarising medium (CX-VIVO + 100 units/mL IL-2 + Th1 polarisation reagents) or maintained in non-polarising control conditions (CX-VIVO + 100 units/mL IL-2) for 5 days (see methods in Section 2.1.8 and Section 4.3.6).

Naïve ZEB2 knockout cells and WT cells polarized to IFN $\gamma$ <sup>+</sup> Th1 cells with similar efficiency (WT; 35%  $\pm$  6.98, and KO; 39%  $\pm$  6.07) (Figure 4.17A) and the quantity of IFN $\gamma$  expressed by the polarised cells, from both the ZEB2 KO and WT populations, was also similar (MFI WT; 3363.86  $\pm$  339.17, and KO; 3516.57  $\pm$  441.4075, Figure 4.17B). The proportion of IFN $\gamma$ <sup>+</sup> cells and IFN $\gamma$  expression in cells grown in non-polarising conditions were much lower compared with Th1 polarising conditions as expected, owing to the lack of differentiation cytokine signalling. However, importantly, the proportion of IFN $\gamma$ <sup>+</sup> cells and IFN $\gamma$  expression in ZEB2 KO compared with WT populations was not significantly different. This important result suggests that ZEB2 may not be involved in Th1 lineage commitment and maintenance, but instead is required for the function of Th1 EM cells.



**Figure 4.16: Gel image of genomic PCR post ZEB2 deletion in Naïve Tconv**  
Post genomic PCR amplicons from WT and KO samples from 3 donors were electrophoresed using 1% agarose and 80V.

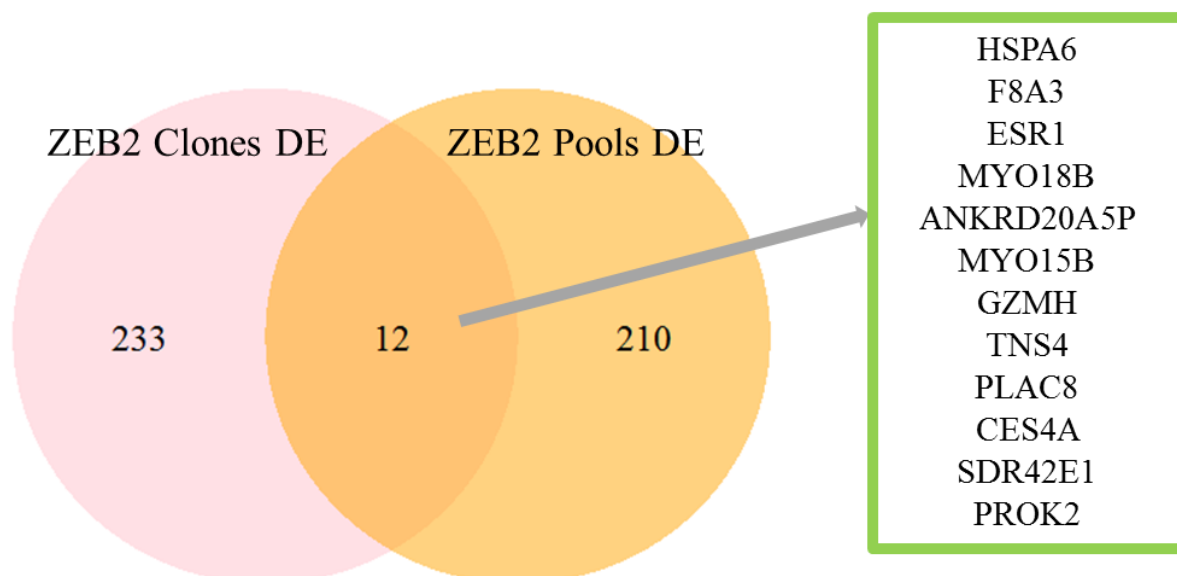


**Figure 4.17: ZEB2 KO Naïve Tconv polarised to Th1**

Human naïve Tconv cells were activated with anti-CD3/anti-CD28 beads for 3 days before Nucleofection™ to delete ZEB2 using triple sgRNAs. Cells were then cultured in Th1 polarising conditions (Th1 cond) or in normal culture medium (Control cond) for 5 days before analysis by flow cytometry. Cells were gated on IFN $\gamma$ + cells for (A) proportion of cells polarised to Th1 and (B) Median Fluorescent Intensity (MFI) of IFN $\gamma$  protein expression. Statistics were carried out with paired t test, not significant (ns) when p value > 0.05, n = 4-7 independent donors.

#### 4.4.5 Common ZEB2 DE between Clones and Pool

Since the different ZEB2 knockout approaches gave very different results, I was interested to investigate if there were any common differentially expressed genes between two RNA-seq datasets. However, there was very little (12 genes) overlap between the two datasets, suggesting that these two different approaches for generation of ZEB2KO result in significant differences in gene expression alterations (Figure 4.18). This has quite important implications for the overall approach to designing these types of experiments. Apart from 4 genes (F8A3, ESR1, ANKRD20A5P, CES4A) most of these common DE genes have opposite expression changes (Figure 4.19). It was not deemed a good use of time and resources to directly compare single vs 3 guide approaches in either pools or clones. However, a clear point of difference was expanding a single cell clone for 2 months compared with growing  $1 \times 10^6$  cells for a few days, which is a very different environment and hence may inadvertently introduce artifacts to the approach.



**Figure 4.18: Overlapping differentially expressed (DE) genes from ZEB2 KO clones and ZEB2 KO pool of Th1 EM cells**

Venn diagram showing overlapped ZEB2 clones DE dataset (pink; generated using single sgRNA) with ZEB2 pools DE dataset (orange; generated using triple sgRNA). Green box displayed 12 genes that are common between the two datasets.



**Figure 4.19: Fold change difference of common genes between ZEB2 KO clones and ZEB2 KO pool of Th1 EM cells dataset**

Bar graphs show the fold change ( $\log_2$ ) of common differentially expressed genes between ZEB2 KO Pools DE RNA-seq dataset (purple; generated using triple sgRNA) with ZEB2 KO Clones DE RNA-seq dataset (orange; generated using single sgRNA). RNA-seq datasets.

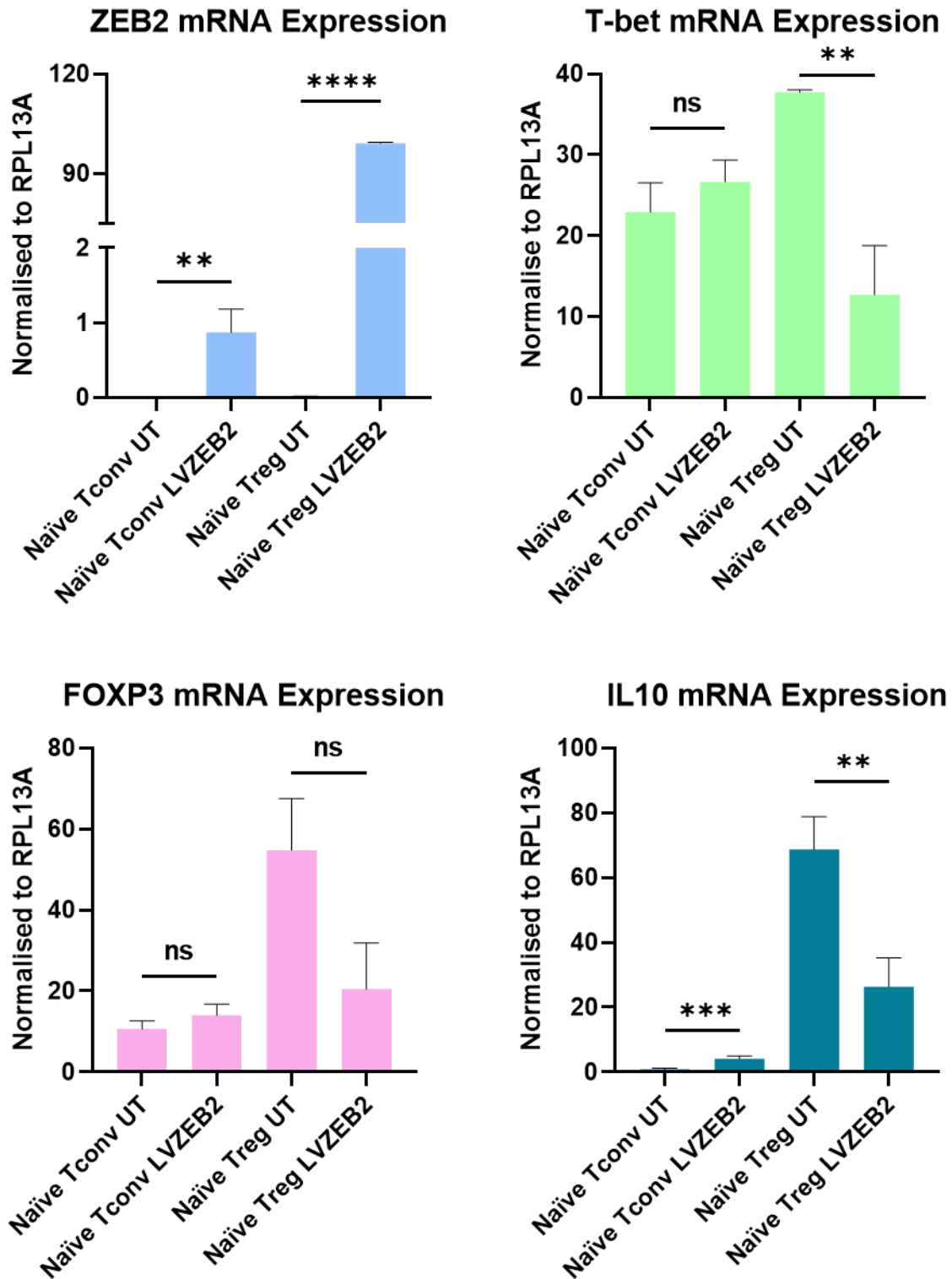
#### 4.4.6 ZEB2 overexpression in Naïve Tconv and Treg

The results from Chapter 3 suggest that ZEB2 a key effector gene involved in promoting the effector function of CD4 Thelper cells, especially Th1 EM. Our lab has shown that ZEB2 expression is tightly controlled in Treg by FOXP3 and FOXP3-induced miR-155 (manuscript in preparation) and this regulatory mechanism maintains the low expression of ZEB2 observed in human Treg compared with its expression in Tconv (Chapter 3: Figure 3.4). It is highly likely that ZEB2 is tightly controlled in Treg by this regulatory mechanism for a good reason. I therefore surmised that perhaps ZEB2 is required to be silent in order to prevent Treg cells from adopting an effector phenotype. However, interesting results in Figure 3.8 showed that ZEB2 is expressed in Treg Effector Memory cells but at a lower level compared with its expression in Tconv Effector Memory cells, suggesting that ZEB2 may possibly have a subtle role in the Treg compartment. Therefore, experiments were designed to study the effect of introducing ZEB2 expression in both naïve Treg and Tconv, where its expression is normally low, as indicated in Figure 3.8.

I first transduced naïve Treg, and also naïve Tconv for comparison, with a lentivirus expressing the ZEB2 open reading frame (ORF); (LV411-ZEB2-IRES-hrGFP, as described in Section 2.5.1). First, I confirmed successful overexpression of ZEB2 in both naïve Tconv and naïve Treg by RT-PCR. ZEB2 mRNA was over ~300 times (Tconv) and ~10000 times (Treg) more highly expressed in the LV-ZEB2-IRES-hrGFP transduced cells than the respective untransduced counterparts (Figure 4.20). I then looked at some selected key genes including FOXP3, T-bet and IL10. Although FOXP3 expression seemed lower in the Treg cells overexpressing ZEB2, this was not significant and so no inferences could be drawn from this observation. However, expression of T-bet was very interesting. In Tconv, where ZEB2 expression is normally high, it was not surprising that overexpressing ZEB2 did not alter T-bet expression. In contrast, overexpression of ZEB2 in Treg had a dramatic effect on T-bet levels.



T-bet is normally highly expressed in Treg, in the Treg1 and Treg1/17 helper lineage population (Chapter 3: Figure 3.14) but overexpression of ZEB2 in Treg markedly reduced expression of T-bet in Treg (Figure 4.20). It would be interesting to speculate that perhaps enforcing expression of ZEB2 in Treg might increase IFN $\gamma$  expression or reduce the ability of Treg to suppress any immune response, in particular, the Th1 immune response. Changes to IL-10 expression with enforced ZEB2 were also interesting. An increase in IL10 expression was observed in Tconv overexpressing ZEB2, suggesting an altered phenotype. IL-10 production by CD4<sup>+</sup> effector T cells has been reported as a mechanism for self-regulation in non-lymphoid tissues (Jankovic et al., 2010). Since Tconv normally express low levels of IL-10, it would be interesting to examine the Th subsets in which IL-10 expression is increased, to determine changes in Tconv function. ZEB2 is normally tightly restricted in Tconv, to the Th1 EM compartment only, and so its aberrant expression elsewhere may be detrimental to the function of other Th subsets. In Treg, where expression of IL-10 is normally high, enforced ZEB2 resulted in dramatically reduced expression of IL-10. IL-10 is the major suppressive cytokine released by Treg and its secretion leads to the suppression of effector immune responses. The results presented here suggest that while ZEB2 is crucial for the effector function of Th1 EM cells, aberrant expression of ZEB2 in both Tconv as well as Treg may impair their normal function. Therefore, tight regulation of ZEB2 is required, not only in Treg where ZEB2 is normally kept silent, but in Tconv as well, where ZEB2 is normally highly restricted to the Th1 EM compartment.



**Figure 4.20: Gene expression changes of ZEB2 overexpression in Naïve Tconv and Treg**  
 Naïve Tconv and naïve Treg cells were pre-activated with anti-CD3/CD28 magnetic coated beads for 1 hour prior to transduction with LV411-IRES-hrGFP (LVZEB2) or without (UT; untransduced). RNA was then collected on day 5 to examine expression of ZEB2, T-bet, FOXP3 and IL10 by qRT-PCR. Relative abundance of ZEB2, T-bet, FOXP3, and IL10 were normalised to reference gene RPL13A and plotted with mean + SEM, statistics were carried out with paired t test, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ,  $n = 3$ .

## 4.5 Discussion

The implications of ZEB2 expression being restricted to Th1 EM are significant, as Th1 are a major effector T cell subtype, producing the Th1 cytokine IFN $\gamma$ , to help promote a Th1 inflammatory response, essential for eliminating viruses and intracellular bacteria. In order to better understand the transcriptional program controlled by ZEB2 in Th1 EM cells, I used CRISPR/Cas9 to delete it in primary human Th1 EM cells, and then explored the consequences. In this chapter, I have demonstrated two methods of deleting ZEB2, one which uses single sgRNAs with clonal expansion and the other was with triple sgRNAs that generate high enough editing efficiency for cells to be used as a pool.

Since my single sgRNA approach used a single cell clone which I then expanded, I noticed that most positively growing clones were not able to grow enough cell numbers for downstream experiments. This could be explained by the findings Kretschmer et al. (2020) that effector memory cells are usually short-lived and that persistence of antigenic but not inflammatory stimuli throughout clonal expansion critically determines its life span. Even though Th1 Effector Memory cells were sort purified, our experimental analyses still revealed heterogeneous behaviour within the pool, and this is very much dependent on the environmental context that each cell experiences. I observed that different clones showed different growth rates and this, in itself, presented difficulties in normalising the harvest period of the clones consistently. These observations may be supported by models of T cell clonal expansion from De Boer et al. (2012) and De Boer and Perelson (2013). Besides, studies by Schumann et al. (2015) suggested that the single sgRNA format is not always efficient. If the sgRNA binds poorly to the desired locus, for instance, Cas9 may fail to cut the DNA. Even when double stranded breaks are produced, some sequences will likely be repaired faithfully. Both of these scenarios may increase the number of wild type sequences in a CRISPR-edited cell pool and thus decrease the frequency of knockout cells as shown in donors #070618 and #150618 (Figure

4.3). Moreover, as there were only 2 donors for this knockout approach experiment and each donor has a limiting clone for one or the other experimental groups (Donor #070615 has 1xWT and 3xKO clones, Donor #150618 has 3xWT and 1xKO clones) this resulted in strong donor bias in the analysis which cannot easily be corrected or removed with RUVg normalisation. Because of this, the differentially expressed genes from RNA-seq data of clones remained questionable for robustness. Therefore, I used a triple sgRNA approach for ZEB2 knockout and this significantly improved upon these shortcomings.

My first major finding using this triple sgRNAs approach for deleting ZEB2 in Th1 EM, was a reduction of IFN $\gamma$  expression. However, this was not observed when polarising ZEB2 knocked out naïve CD4<sup>+</sup> T cells to a Th1. Maybe ZEB2 has a very specific role in already differentiated Th1. As this is a signature Th1 effector gene, this suggests a role for ZEB2 in Th1 lineage fidelity. Mechanistically, this could be by direct or indirect regulation of the IFN $\gamma$  gene. Slade et al. (2020) has recently reported that PLAC8 was important in suppressing IFN $\gamma$  mRNA and protein production. Interestingly, PLAC8 was upregulated when ZEB2 is knocked out, suggesting a possibility that the reduction of IFN $\gamma$  was owing to the upregulation of PLAC8 which may be directly regulated by ZEB2.

Additionally, deleting ZEB2 resulting in increasing GZMA and GZMK mRNA expression. Intracellular staining of these granzymes protein may further provide indication that the Th1 EM may become more actively armed to release cytotoxic effector molecules owing to the loss of ZEB2. Notably, ZEB2 has previously been associated with cytotoxic CD4 (CD4-CTL) function based on similarities with CD8-CTL. The coincidence of CTL-like activity in CD4 helper subsets remains controversial, and interrogation of scRNA-seq datasets (Patil et al., 2018, Meckiff et al., 2020, Uniken Venema et al., 2019) weakly associates ZEB2 as a co expressed gene in the cytotoxic CD4 clusters during infection. This is plausible, as Th1 are poised to eliminate virally infected cells, and this can be by cytolysis. My strategy of using the

chemokine receptor, CXCR3, in conjunction with CD62L to purify Th1 EM might enrich for CD4-CTL. Therefore, analysis of RNA-seq between Th1 CM and Th1 EM could be carried out to look for enrichment of key CD4-CTL signature genes such as CRTAM, EOMES, PRF1, GNLY and GZMB. Staining of CRTAM protein in Th1 EM cells could also verify if they are CD4-CTL. Overall, my data suggest that ZEB2 is required to maintain the function of Th1 EM and restrict CTL function.

I also observed downregulation of costimulatory receptors mRNA expression such as CD27 and CD28 on the ZEB2 KO cells, which are known markers for lymphocytes with cytotoxic activity. However these molecules do not necessarily represent authentic markers for CD4 CTLs (Brown, 2010). Since deletion of ZEB2 resulted in an increased expression of several central memory signature genes including CD27 and SELL (SELL encodes L-selectin aka CD62L) and decreases in several effector memory signature genes compared with the WT, this suggests the KO cells are partially reverting into a CM phenotype or may be regressing to an early EM (CD62L<sup>hi</sup> CD27<sup>+</sup>) phenotype from a late EM phenotype (CD62L<sup>lo</sup> CD27<sup>-</sup>) based on phenotypes reported by Stephens and Langhorne (2010). It could also be that ZEB2 does not necessarily prevent cells from reverting to CM but instead it is the functional properties of SELL (CD62L, L-selectin) that ZEB2 regulates. Increased L-selectin is a homing receptor facilitating migration to secondary lymph organs rather than carrying out effector functions. Besides being a memory marker, CD27 also promotes Th1 formation and survival and is implicated in decreasing Th1 stability if over-expressed, as this disrupts normal T-cell homeostasis leading to hyperactivation (van Oosterwijk et al., 2007). I postulate that ZEB2 might directly or indirectly regulate both CD27 and SELL by transcriptional repression, so the loss of ZEB2 in Th1 EM in disease may cause cells to revert to a precursor effector memory phenotype enabling it to recirculate in nearby tissues, and with the ability to induce cytotoxic molecules, to cause tissue damage.

Overall, the results of this chapter support the hypothesis that ZEB2 is required to maintain Th1 function and fidelity. It is also worth mentioning that even though the triple sgRNA approach for ZEB2 deletion was more robust than the single sgRNA approach, further optimisation to further increase the INDEL efficiency to around ~90% could be carried out. These further optimisation strategies could include testing out other sgRNA sequences. The CRISPR/Cas9 knockout field would also benefit from having a reliable marker to sort purified edited cells from RNPs Nucleofection™ as suggested by van Leeuwe et al. (2019). In the effector arm of CD4<sup>+</sup> T cells, ZEB2 plays a pivotal role in ensuring that the Th1 EM cells mount a correct Th1 response by maintaining IFN $\gamma$ . However, it is possible that when a certain threshold level of ZEB2 expression is reached, ZEB2 could potentially induce IL-10 cytokine expression and trigger self-regulating mechanisms after foreign pathogen clearance. In the suppressor arm of CD4<sup>+</sup> T cells, ZEB2 may also potentially get turned on in a Treg, as I observed in Treg Effector Memory cells, (previous chapter). High ZEB2 expression in a Treg was shown to repress IL-10 expression, which may potentially prevent the Effector Memory Treg from carrying out its regulatory functions while they are recirculating between peripheral blood and non-lymphoid tissues. However, if suppressive activity is required, FOXP3 may tightly repress ZEB2 so that IL-10 can be secreted to modulate the immune response. Loss of IL-10, by enforced expression of ZEB2, would be highly likely to impair the function of Treg, which, if time permitted, could be shown by the ability of the Treg to suppress T effector proliferation. ZEB2 does get turned on in a Treg, as seen in Treg Effector Memory cells, presumably by other mechanisms (as mentioned in the previous chapter). Given the well-established role of ZEB2 in cancer metastasis, perhaps ZEB2 facilitates Treg migration? The high ZEB2 expression may potentially be there to keep the Treg Effector Memory from carrying out its suppressive or regulatory function while they were recirculating between peripheral blood and non-lymphoid

tissues. However, if suppression activity is required, FOXP3 may tightly repress ZEB2 so that IL-10 can be secreted to modulate the immune response

All of these suggest that ZEB2 most likely helps to maintain the lineage fidelity of Th1 EM cells and thus has important implications in autoimmune disease. Interestingly, overexpression of ZEB2 showed increased expression of the anti-inflammatory cytokine IL10 in naive Tconv, and conversely reduced IL10 expression in naive Treg, suggesting a differential role of ZEB2 in the Tconv and Treg compartment. In a normal healthy immune response, ZEB2 may play a pivotal role in ensuring that Th1 EM cells of the effector arm (Tconv) mount a correct Th1 response. It is believed that the Th1 response is involved in the pathogenesis of multiple autoimmune diseases such as Type 1 diabetes mellitus, Rheumatoid arthritis, Systemic Lupus Erythematosus (SLE) and Inflammatory Bowel Disease (IBD) and so it is crucially important to have tight regulation of pro-inflammatory immune responses. My novel findings elucidate the importance of ZEB2 in Th1 Effector Memory cells and additionally, suggest that ZEB2 may have a role in the Treg compartment and if that is dysregulated as part of these diseases, it could therefore be a therapeutic target to mitigate or prevent immune dysregulation.

## **CHAPTER 5: META-ANALYSIS OF ZEB2 RNA-SEQ DATA**



## 5.1 Introduction

The results in chapter 3 and 4 suggest that ZEB2 is required for correct Th1 cell immune function; orchestrating a specific transcriptional repertoire to help the Th1 elicit and maintain its specific pro-inflammatory function. ZEB2 clearly influences the expression of many genes (described in Chapter 4) either directly, or indirectly, via a network of interactions and signalling pathways, but my analysis so far has yielded little detailed information about the nature and complexity of these interactions. Unravelling the relationships between these genes to fully reveal the role of ZEB2 in a Th1 cell, requires the more detailed analytical approach using bioinformatics.

Here, I have used a set of computational approaches to carefully and fully reveal the network of interactions orchestrated by ZEB2. Pathway enrichment was carried out to provide mechanistic insight into the roles of the differentially expressed genes that I have obtained from my ZEB2 knockout studies (described in Chapter 4). This method uses statistical parameters to identify biological pathways that are enriched in a gene list resulting from RNA-seq and genome-sequencing experiments more than would be expected by chance. The protocol comprises three major steps: definition of a gene list from transcriptomic data, determination of statistically enriched pathways, and visualization and interpretation of the results. The resulting method of analysis enables discovery of novel biological functions, genotype-phenotype relationships, and disease mechanisms.

Gene Ontology (GO) analysis provides a system for hierarchically classifying genes or gene products into terms organized in a graph structure (or an ontology). The terms are grouped into three broad categories: molecular function (describing the molecular activity of a gene), biological process (describing the larger cellular or physiological role carried out by the gene, coordinated with other genes) and cellular component (describing the location in the cell where the gene product executes its function). My RNA-seq data from Chapter 4 comprises sets of

genes that are differentially expressed. The GO term gene set can be used to functionally profile this set of genes, to determine which GO terms appear more frequently than would be expected by chance when examining the set of terms annotated to the input genes. KEGG is a database resource for understanding high-level function and utility of the biological system, such as the cell, from molecular-level information, especially from large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.

Genome-wide association studies (GWAS) provides a means of identifying genes/ regions of the genome that contain genetic risk that is associated with a trait such as human diseases. This method searches the genome for small variations, called single nucleotide polymorphisms, or SNPs, that occur more frequently in people with a particular disease than in people without the disease. Enrichr, is an integrative web-based interface that was used to rank the significance of disease associated GWAS terms (GWAS catalogue 2019) using the list of differentially expressed genes from the deletion of ZEB2 in Th1 EM, by comparing the Fisher exact test to a method it uses which then computes the deviation from the expected rank for terms.

ZEB2 has a well-documented role in promoting epithelial-to-mesenchymal transition (EMT), transcriptionally repressing genes via binding to bipartite E-box motifs in gene regulatory regions (Sekido et al., 1994). Despite the abundant presence of E-box motifs within the human genome and the multiplicity of pathophysiological processes regulated during ZEB2-induced EMT, only a small fraction of validated ZEB2 targets have been identified so far. Hence, we explored the public domain for mineable datasets including genome-wide ZEB2 binding by chromatin immunoprecipitation-sequencing (Davis et al., 2018) under endogenous ZEB2 expression conditions, which can be filtered by chromatin accessibility (ATAC-seq; (Calderon et al., 2019)) in a Th1 cell.

Transcriptional profiling is an effective approach to provide biological insight and rapid, unbiased screening of whole transcriptomes to reveal the most promising genes or markers associated with the function of ZEB2 in Th1 Effector Memory cells. However, the previous studies were mostly concerned with differentially expressed (DE) genes individually and did not consider clusters of highly correlated genes, which may be responsible for specific clinical features of interest. Therefore, in this chapter, I will identify and analyse co-expression modules within the Th1 EM WT and ZEB2 KO pool RNA-seq datasets, using a statistical tool in the analytical program R known as Co-Expression Modules identification Tool (CEMiTool) (Russo et al., 2018). It allows for an unsupervised gene filtering method, automated parameter selection for identifying modules, enrichment and module functional analyses, as well as integration with interactome data. Therefore, providing a novel insight into which functional regulatory groups of genes may be driving the differences in transcriptional signatures in Th1 EM cell function owing to the loss of ZEB2.

The computational and enrichment analysis carried out in this chapter indicates that the ablation of ZEB2 shows enrichment of signatures associated with multiple autoimmune diseases including Inflammatory Bowel Disease (IBD) from both KEGG and GWAS analysis. Using the published ZEB2 ChIP-seq data, the DE genes that were a plausible direct target of ZEB2 were also investigated. Interestingly, ZEB2 deleted Th1 Effector Memory cells also showed increase expression of genes usually found in a Th1 Central Memory cell type. The results from this work suggest the importance of ZEB2 in multiple aspects of the Th1 cell biology, to maintain immune homeostasis and prevent the pathogenesis of autoimmune diseases such as IBD.

## 5.2 Aims and Hypothesis

The hypothesis for this chapter is that dysregulation of ZEB2 in Th1 may be linked to the pathogenesis of IBD

The focus of this chapter is to address Aim 3 of my PhD proposal.

- 3.1. To perform enrichment analysis on the DE genes from ZEB2 KO in both clones and pools of Th1 EM cells KO.
- 3.2. To determine DE genes that are directly regulated by ZEB2.
- 3.3. To compare Th1 CM vs Th1 EM with ZEB2-KO Th1 EM vs Th1 EM.
- 3.4. To perform comprehensive modular co-expression analyses.

## 5.3 Material & Methods

### 5.3.1 Gene Ontology (GO) Functional Annotation Analyses

To consider the higher order biological processes, molecular function and cellular component from differential expression results, one of the most commonly used resources is Gene Ontology (GO) databases, which annotate genes according to a dictionary of hierarchical nested annotation terms. GO terms that occur frequently in the list of DE genes are over-represented or enriched. In R package limma, GO analyses was conducted using the goana() function on the output of the differential expression analysis (Section 2.4.6) comparing WT and KO of either Th1 EM clones or Th1 EM pools. The goana function was used on the DE genes to conduct overlap tests for the up- and down-regulated DE genes separately. An FDR cutoff of 5% or 10 % was used when extracting DE genes, but this can be varied. As goana relies on entrezgene identifiers, analysis was switched to these IDs for this step of the analysis. Results were then limited to GO terms with at least 4 steps back to the root terms for each term to remove of high-level GO terms, which resulted in the omission of less informative or redundant GO terms from the analysis. In order to visualise the results of the GO analysis, the R package clusterProfiler was used to plot the Gene-Concept Network plot known as cnetplot.

### 5.3.2 KEGG Pathway Functional Annotation Analyses

Another robust annotation database is the Kyoto Encyclopedia of Genes and Genomes (KEGG). Much smaller than GO, this is a curated database of molecular pathways and disease signatures. A KEGG analysis was carried out exactly as for GO in Section 5.3.1, but using the kegg() function from the R package limma. By default, the kegg() function automatically reads the latest KEGG annotation update from the originators each time it is run.

### 5.3.3 Gene Set Enrichment Analysis

There are numerous methods for testing within a ranked list, with the most widely used and most well-known being Gene Set Enrichment Analysis (GSEA). This was performed natively

in R using a package known as fgsea which is a fast implementation of the GSEA algorithm. This is also wrapped by the function GSEA() from the package clusterProfiler, which can later be used for visualisation. To perform GSEA, all genes from Section 2.4.6 were ranked by  $-\log_{10}p$ , multiplied by the sign (+/-) of the logFC. In this convention, up-regulated genes will receive a positive score, whilst down-regulated genes will receive a negative score. Hallmark gene sets were used for enrichment of specific well-defined biological states, or processes, that display coherent expression. The hallmark gene set files for this analysis were obtained from the GSEA website ([www.broadinstitute.org/gsea/](http://www.broadinstitute.org/gsea/)). Enrichment scores (ES) and false discovery rate (FDR) values were applied to sort genes after gene set permutations were performed 100000 times for the analysis. Barcode plots were then used for visualisation of the GSEA results.

### 5.3.4 Enrichr GWAS 2019 Analysis

Analysis for GWAS conducted with Enrichr (Kuleshov et al., 2016). Sets of significantly upregulated or downregulated ( $FDR < 0.05$  or  $FDR < 0.1$ ) genes were separately queried to test the likelihood if certain GWAS terms, from GWAS catalogue 2019, were statistically overrepresented among differentially expressed genes. Enrichr will then calculate p-values, adjust p-values, Z-scores, and combines representative scores of Fisher exact and Z-score statistics. For all analyses via Enrichr, enrichments were considered statistically significant if the p-values was less than 0.05.

### 5.3.5 ZEB2 ChIP Filtered by ATACseq

#### Global ZEB2 ChIP-seq

ZEB2 chromatin immunoprecipitation-sequencing (ChIP-seq) datasets from K562 and HEK293 available from the ENCODE project (Davis et al., 2018) (<https://www.encodeproject.org/>) with the following accession identifiers: ENCSR004GKA (K562), ENCSR322CFO (K562), ENCSR417VWF (HEK293). The normalised bed files

(narrowpeak) from three ChIP-seq data files were then merged together to account for all regions in which ZEB2 could potentially bind. These regions will later be filtered on chromatin accessibility of an activated Th1 cell.

### Th1 ATAC-seq

Chromatin accessibility from activated Th1 ATAC-seq data was acquired from published studies (Calderon et al., 2019). Briefly, the adapter sequences were removed from the raw fastq data using AdapterRemoval (v2.2.1). Paired-end ATAC-seq reads were then aligned to the human genome (GRCh38.98, downloaded from ENSEMBL), using bowtie2 (v2.4.2) (parameters: -k 10 --very-sensitive). First, PCR duplicates were removed using Picard (v1.119). Reads mapped to mitochondrial DNA were then excluded. Only uniquely mapped and properly paired reads with insert size <2 kb and mapping quality over 30 were kept for downstream analysis. ATAC-seq peak calling was performed with Genrich (v0.6, available at <https://github.com/jsh58/Genrich>, parameters: -r -m 30 -q 0.05 -a 200 -j -y -e MT,Y -b) for each region. ATAC-seq data of the same region from different individuals were treated as replicates in peak calling, to generate peak sets for each region.

### Annotation and Intersection

The combined ZEB2 ChIP data (described above) were then intersected with Th1 ATAC-seq data using bedtools (v2.30.0, parameters: -wa -u) to retain ChIP peaks that have overlap with chromatin accessibility in activated Th1 cells. The filtered ZEB2 ChIP peaks were then annotated to the nearest gene (upstream: 5000, downstream: 1000) using a tool for Peak Annotation and Visualization (PAVIS; <https://manticore.niehs.nih.gov/pavis2>) (Huang et al., 2013). To identify genes in the DE gene list from ZEB2 KO in Th1 EM (Section 4.4.3) that are a target of the ZEB2 transcription factor, annotated genes from ZEB2 ChIP filtered by activated Th1 chromatin accessibility were then overlapped with the DE gene list to look for common genes. These genes were identified as a direct target of ZEB2.

### 5.3.6 Gene Co-expression Network Analysis with CEMiTool

For the construction of a weighted and signed co-expression network, I used the R (version 4) package CEMiTool (version 1.9.3) (Russo et al., 2018). This offers an implemented unsupervised gene filtering method and a more automated parameter selection, aiming to enhance the reproducibility of results. To reduce noise, the filtering procedure starts with retaining 6000 genes with the most variable gene expression across samples. Then, the variance of genes is modelled as an inverse gamma distribution, and genes are filtered based on a p-value. The standard p-value of 0.1 was chosen as this was shown to be a good compromise between noise reduction and information loss. To construct the network, Pearson correlation was calculated as a similarity measure for all pair-wise genes. In CEMiTool, the  $\beta$  parameter is selected by an algorithm, which is based on the concept of Cauchy sequences. For adherence to the scale-free topology, only  $\beta$  values with  $R^2 > 0.80$  are considered, while lower  $\beta$  values are preferred due to considerations of network connectivity. The default parameters in CEMiTool were contained to preserve the advantages of more reliable and consistent results with the automatic selection of  $\beta$ . Clusters of highly co-expressed genes (modules) were detected by a dynamic algorithm for selecting branches of the hierarchical clustering dendrogram implemented in the Dynamic Tree Cut package. The minimum number of genes per submodule was set to the default of 30 genes and the module merging correlation threshold for eigengene similarity was 0.8. To determine biological functions associated with the modules, the Hallmark gene set list from MSigDB was used to perform an over representation analysis via the clusterProfiler R package in CEMiTool. Top gene sets enriched on each co-expression module were detected by the hypergeometric test with  $p = 0.05$ . To visualize interactions between the genes in each co-expression module, the combined human interaction data from GeneMania was included in the analysis. In order to assess differences in module activity between the early adversity and control group, a gene set enrichment analysis with the



fgsea R package was performed with the default values in CEMiTool. Here, genes from co-expression modules were used as gene sets and the z-score normalized expression of the samples within each group are ranked in the analysis. The enrichment score (ES) is normalized by taking into account the size of each gene set, leading to a normalized ES (NES). The proportion of false positives is controlled by calculating the false discovery rate corresponding to each NES and an adjustment of the respective p-value.

## 5.4 Results

### 5.4.1 Enrichment Analysis for ZEB2 KO Th1 EM Clones

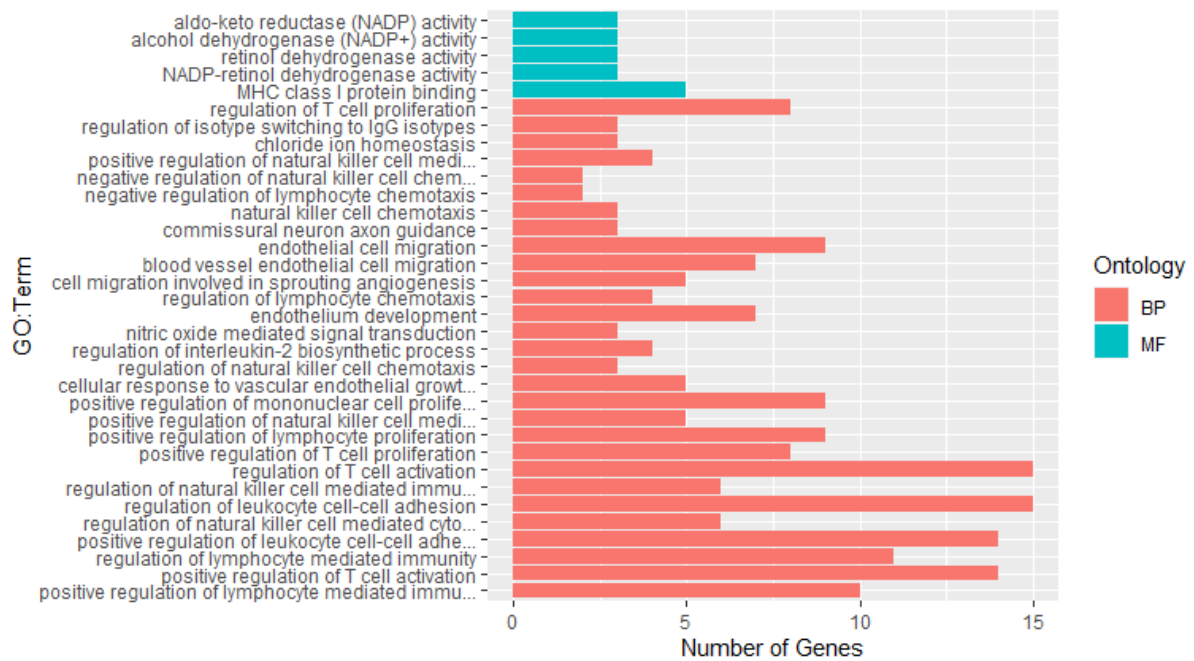
In Chapter 4, I demonstrated that the knockout of ZEB2 in both Th1 EM cell clones (single sgRNA knockout approach; Section 4.4.2) and Th1 EM pools (triple sgRNA knockout approach; Section 4.4.3) resulted in multiple genes becoming differentially expressed. However, the functional significance of these differentially expressed genes was not defined. In order to provide mechanistic insight into the biological outcome of genes affected by the loss of ZEB2, the following enrichment analyses were performed.

#### Gene Ontology annotation

I first performed GO enrichment analysis on my RNA-seq data from Th1 EM cell clones (generated from Section 4.4.2), using the default settings of the function `goana()` in R, on the differentially expressed genes selected based on  $FDR < 0.05$ . As the `goana()` function relies on entrezgene identifiers (EntrezID), analysis was switched to these IDs for this step. The input number of genes for this analysis was 217 instead of the initial 245 as some genes did not have an EntrezID. Results were then limited to GO terms with at least 4 steps back the root terms for each term.

A total of 34 GO terms ( $FDR < 0.05$ ) were identified amongst the differentially expressed genes, 29 of which were from the Biological Process term and 5 were from the Molecular Function term (Figure 5.1). Most of the Biological Process GO terms comprised positive regulation of T cell activation, cell-cell adhesion and proliferation (Table 5.1). Interestingly, there were many enriched terms associated with Natural Killer cell function. This observation might suggest that the loss of ZEB2 allows the Th1 cells to acquire characteristics of Natural Killer cells. The molecular function GO terms that were statistically enriched were mostly related to NADP<sup>+</sup>/NADP activity which may result in oxidative stress affecting the Th1/Th2 ratio (Table 5.1) owing to the loss of ZEB2. Interestingly, MHC class I binding was enriched in WT Th1

EM clones even though Th1 cells do not normally interact with MHC class I molecules. This could be explained by the possibility that these WT clones acquired expression of CD8A co-receptor during long term clonal expansion culture and allowed them to facilitate antigen recognition of an MHC class I-restricted response. Studies by Kessels et al. (2006) and Willemsen et al. (2005) have previously demonstrated the possibility of redirecting CD4<sup>+</sup> T cells towards antigens presented by MHC class I through the introduction of the CD8A gene. However, ZEB2 deleted Th1 EM clones have reduced expression of CD244, CD8A and PILRA genes (involved in MHC class I protein binding) suggesting less capability in recognition of antigens presented by MHC class I in comparison to WT clones. Some of the Natural Killer cell-mediated cytotoxicity and immunity Biological Process terms were also enriched in WT Th1 EM clones and this could also be owing to the long-term culture resulting in the reprogramming of Th1 to Natural Killer -like cells which do have the ability to interact with MHC class I molecules as well. Genes significantly associated with these GO terms are also visualized in Figure 5.2 for Biological Process terms and Figure 5.3 for Molecular Function terms. Not surprisingly, some genes were interconnected between two or more enriched GO terms.

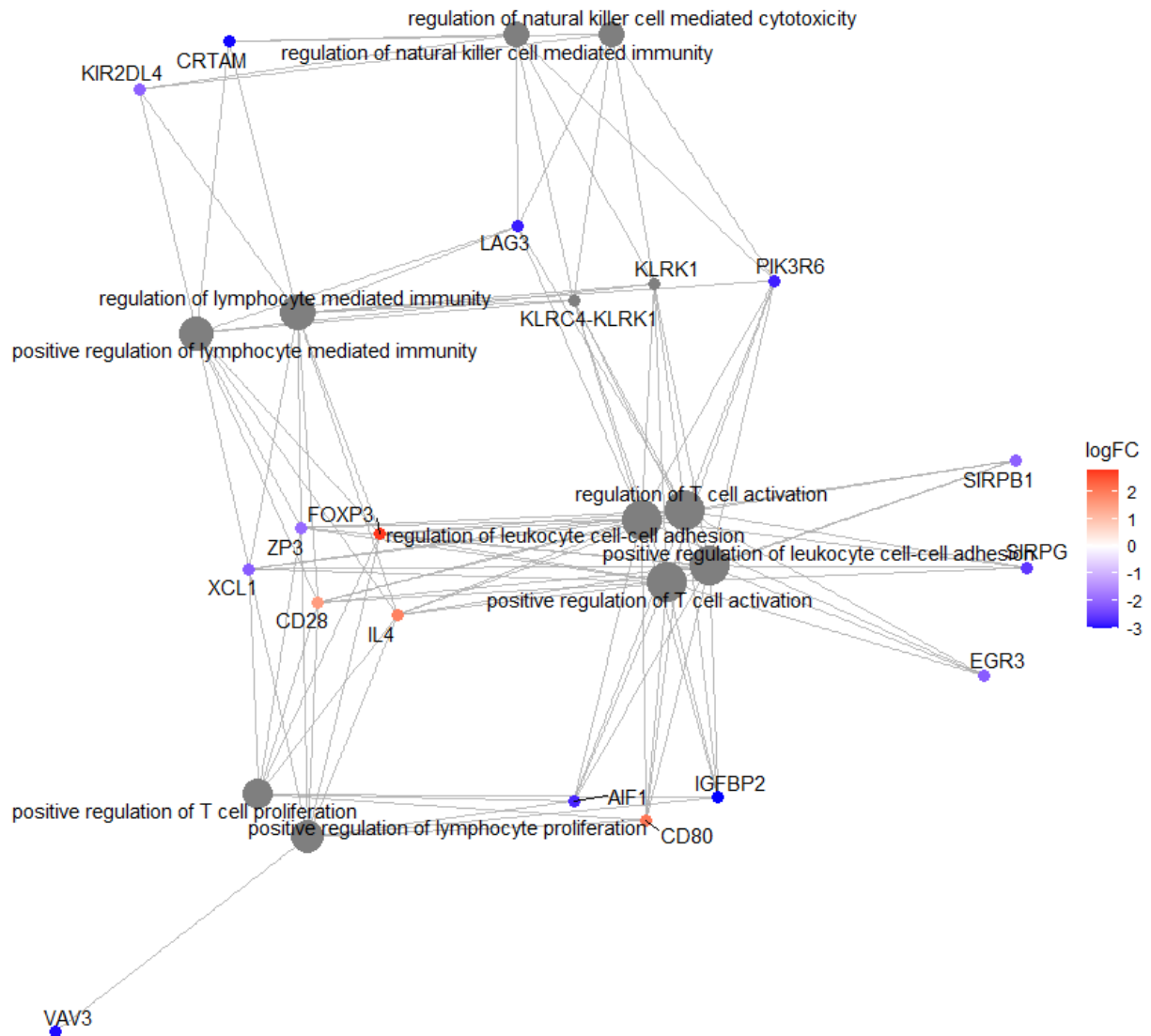


**Figure 5.1: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM clones**

Biological Process (BP) and Molecular Function (MF) were annotated as part of the GO terms.

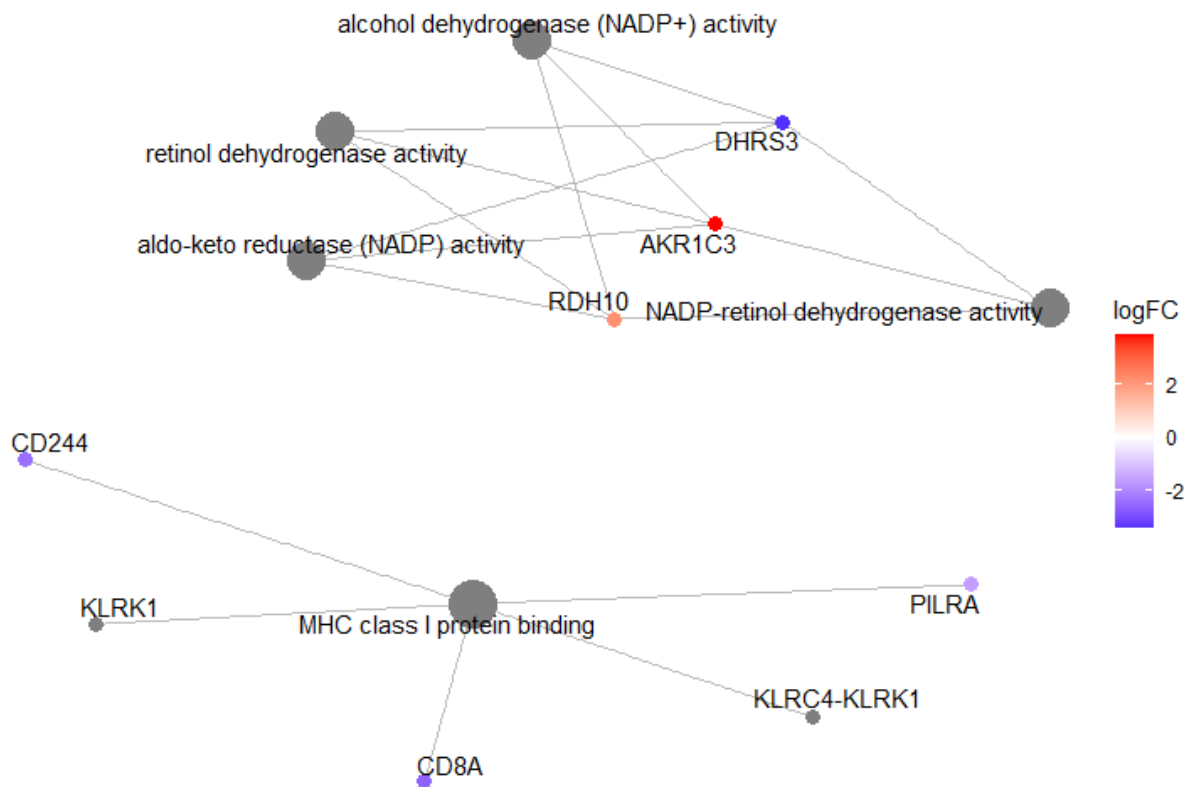
**Table 5.1: Significantly enriched GO terms using an FDR < 0.05 and using ZEB2 KO Th1 EM clones DE genes with FDR < 0.05**

GO ID	Term	Ontology	N	Obs DE	Expected DE	FDR
GO:0002708	positive regulation of lymphocyte mediated immunity	BP	74	10	1.4	1.53E-03
GO:0050870	positive regulation of T cell activation	BP	155	14	3	1.53E-03
GO:0002706	regulation of lymphocyte mediated immunity	BP	100	11	1.9	1.73E-03
GO:1903039	positive regulation of leukocyte cell-cell adhesion	BP	166	14	3.2	1.73E-03
GO:0042269	regulation of natural killer cell mediated cytotoxicity	BP	29	6	0.6	4.91E-03
GO:1903037	regulation of leukocyte cell-cell adhesion	BP	215	15	4.1	4.91E-03
GO:0002715	regulation of natural killer cell mediated immunity	BP	31	6	0.6	5.66E-03
GO:0050863	regulation of T cell activation	BP	230	15	4.4	7.74E-03
GO:0042102	positive regulation of T cell proliferation	BP	68	8	1.3	8.68E-03
GO:0050671	positive regulation of lymphocyte proliferation	BP	91	9	1.7	9.79E-03
GO:0002717	positive regulation of natural killer cell mediated immunity	BP	23	5	0.4	9.79E-03
GO:0032946	positive regulation of mononuclear cell proliferation	BP	92	9	1.8	9.79E-03
GO:0035924	cellular response to vascular endothelial growth factor stimulus	BP	25	5	0.5	1.36E-02
GO:2000501	regulation of natural killer cell chemotaxis	BP	6	3	0.1	1.78E-02
GO:0045076	regulation of interleukin-2 biosynthetic process	BP	15	4	0.3	1.90E-02
GO:0007263	nitric oxide mediated signal transduction	BP	7	3	0.1	2.31E-02
GO:0003158	endothelium development	BP	65	7	1.2	2.31E-02
GO:1901623	regulation of lymphocyte chemotaxis	BP	17	4	0.3	2.43E-02
GO:0002042	cell migration involved in sprouting angiogenesis	BP	31	5	0.6	2.43E-02
GO:0043534	blood vessel endothelial cell migration	BP	67	7	1.3	2.43E-02
GO:0043542	endothelial cell migration	BP	114	9	2.2	2.63E-02
GO:0071679	commissural neuron axon guidance	BP	8	3	0.2	2.63E-02
GO:0035747	natural killer cell chemotaxis	BP	8	3	0.2	2.63E-02
GO:1901624	negative regulation of lymphocyte chemotaxis	BP	2	2	0	2.63E-02
GO:2000502	negative regulation of natural killer cell chemotaxis	BP	2	2	0	2.63E-02
GO:0045954	positive regulation of natural killer cell mediated cytotoxicity	BP	20	4	0.4	3.44E-02
GO:0055064	chloride ion homeostasis	BP	9	3	0.2	3.47E-02
GO:0048302	regulation of isotype switching to IgG isotypes	BP	10	3	0.2	4.59E-02
GO:0042129	regulation of T cell proliferation	BP	103	8	2	4.65E-02
GO:0042288	MHC class I protein binding	MF	18	5	0.3	4.91E-03
GO:0052650	NADP-retinol dehydrogenase activity	MF	7	3	0.1	2.31E-02
GO:0004745	retinol dehydrogenase activity	MF	7	3	0.1	2.31E-02
GO:0008106	alcohol dehydrogenase (NADP+) activity	MF	9	3	0.2	3.47E-02
GO:0004033	aldo-keto reductase (NADP) activity	MF	10	3	0.2	4.59E-02



**Figure 5.2: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM clones**

CNET plot showing top 10 significantly enriched Biological Process GO terms (grey hub modules) using an FDR < 0.05 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM clones with FDR < 0.05.



**Figure 5.3: Gene-concept network plot showing all Molecular Function GO terms from Th1 EM clones**

CNET plot showing all significantly enriched Molecular Function GO terms (grey hub modules) using an FDR < 0.05 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM clones with FDR < 0.05.

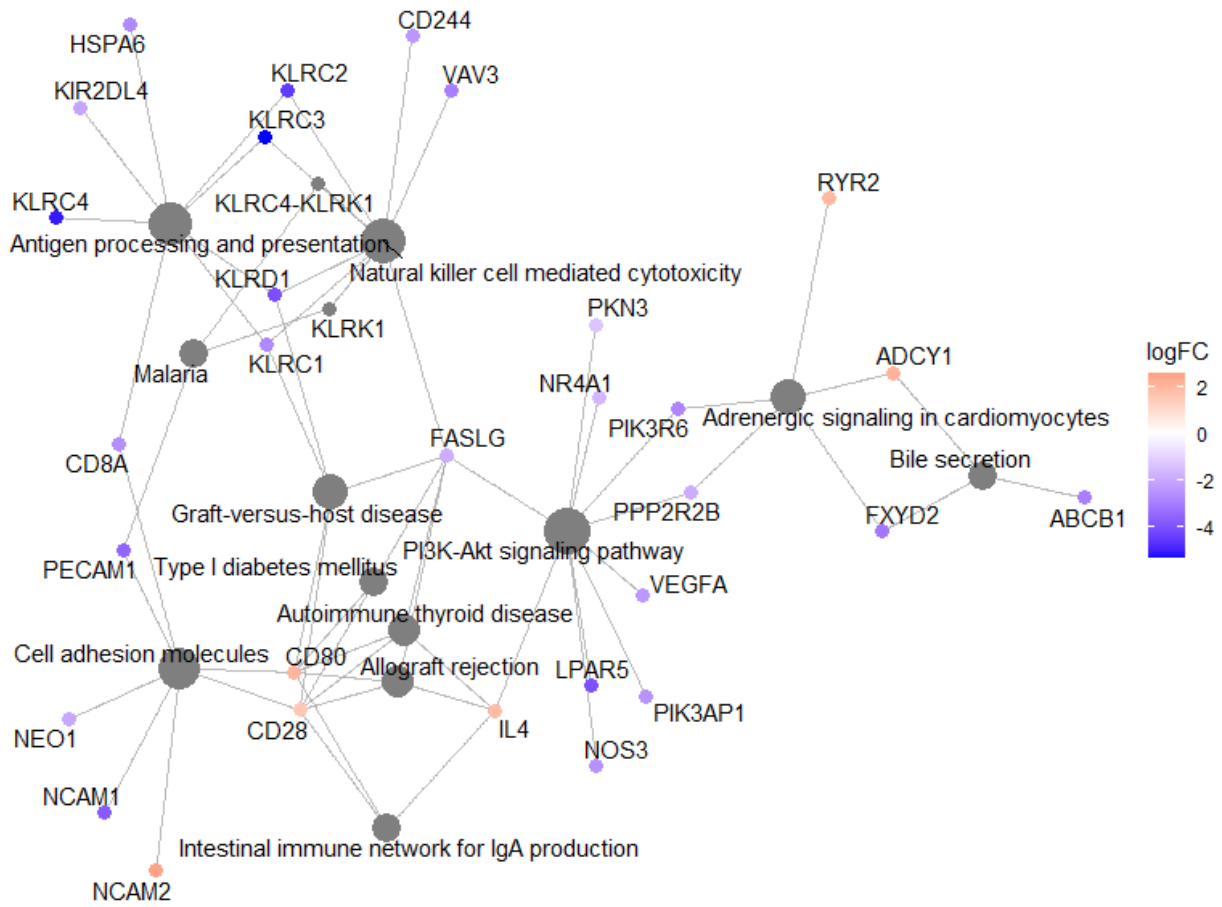
[KEGG pathway annotation](#)

An analysis for detection of enriched KEGG terms was also performed. KEGG pathway enrichment analysis showed that, among the 17 enriched pathways based on  $p$ -value  $< 0.05$ , antigen processing and presentation, natural killer cell mediated cytotoxicity, cell adhesion, immune related diseases and pathways in cancer were significantly enriched (Table 5.2). Affected genes in those KEGG pathways were also shown as a CNETplot in Figure 5.4. This analysis also showed the enrichment of Natural Killer cell mediated cytotoxicity as shown in previous GO term annotation. Based on the KEGG pathway analysis, it was suggested that ZEB2 has a role in cell adhesion, T cell activation via antigen processing and presentation, as well as immune related diseases via dysregulation of multiple genes. The connection between differentially expressed genes with fold changes and aforementioned KEGG terms was also mapped by CNETplot to show a network of genes affected using those terms. Owing to the variability of Th1 EM clones as discussed in Chapter 4, I determined that these results were not entirely reliable and therefore not discussed any further. Next, I repeated the experiments using an optimised protocol of the three sgRNA guide system for CRISPR knockout of ZEB2 and pools of cells, rather than Th1 EM clones.

**Table 5.2: Significantly enriched KEGG terms using a  $p$ -value  $< 0.05$  and using ZEB2 KO Th1 EM clones DE genes with FDR  $< 0.05$**

KEGG ID	Pathway	N	Obs DE	Expected DE	P-value
hsa04612	Antigen processing and presentation	62	8	1.2	2.23E-05
hsa04650	Natural killer cell mediated cytotoxicity	87	9	1.7	4.16E-05
hsa05332	Graft-versus-host disease	29	5	0.6	2.02E-04
hsa04514	Cell adhesion molecules	79	7	1.5	7.78E-04
hsa05320	Autoimmune thyroid disease	29	4	0.6	2.14E-03
hsa05330	Allograft rejection	30	4	0.6	2.44E-03
hsa04151	PI3K-Akt signaling pathway	191	10	3.7	3.76E-03
hsa05144	Malaria	19	3	0.4	5.37E-03
hsa04976	Bile secretion	28	3	0.5	1.60E-02
hsa04940	Type I diabetes mellitus	31	3	0.6	2.11E-02
hsa04672	Intestinal immune network for IgA production	32	3	0.6	2.29E-02
hsa04261	Adrenergic signaling in cardiomyocytes	86	5	1.6	2.46E-02
hsa05200	Pathways in cancer	329	12	6.3	2.48E-02
hsa00830	Retinol metabolism	14	2	0.3	2.86E-02
hsa05418	Fluid shear stress and atherosclerosis	91	5	1.7	3.05E-02
hsa03320	PPAR signaling pathway	36	3	0.7	3.12E-02
hsa00770	Pantothenate and CoA biosynthesis	15	2	0.3	3.26E-02
hsa00140	Steroid hormone biosynthesis	16	2	0.3	3.68E-02
hsa04911	Insulin secretion	42	3	0.8	4.62E-02





**Figure 5.4: Gene-concept network plot showing top 12 significantly enriched KEGG pathway terms from Th1 EM clones**

CNET plot showing top 12 significantly enriched KEGG pathway terms (grey hub modules) using a p-value < 0.05 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM clones with FDR < 0.05.

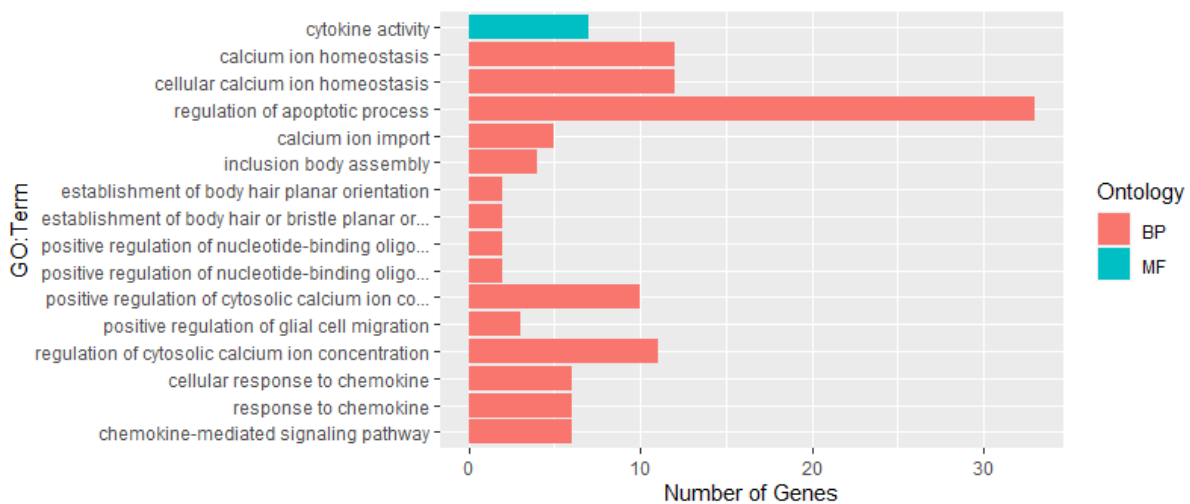
### 5.4.2 Enrichment Analysis for ZEB2 KO Th1 EM Pools

The Enrichment analysis described above, was performed on the DE gene list from the ZEB2 knockout in the Th1 EM clones (Section 5.4.1). As explained above, due to concerns about long term clonal expansion causing high variance and phenotypic instability, I decided to repeat these experiments using a 3 sgRNA guide CRISPR system in Th1 EM cell pools. The same analysis as before, was carried on the DE gene list from the ZEB2 knockout Th1 EM pools, but using two different False Discovery Rate (FDR) stringency strategies:  $FDR < 0.05$  or  $FDR < 0.1$ . The FDR determines whether a differentially expressed gene is statistically significant based on the threshold value. The approach of performing enrichment analysis with two FDR cut-off allows for detection of more subtle changes and accounts for the KO efficiency in the pooled Th1 EM cells. Whilst no significant effects would be expected on the top-ranked DE genes, it is conceivable that this may impact the results following enrichment analysis and may reveal more subtle changes. Again, enrichment analysis was performed with EntrezID instead of the gene symbols. Since some of the genes do not have an EntrezID, these genes were excluded. Therefore, the input number of genes using gene list at  $FDR < 0.05$  was 205 rather than the initial 222, and for  $FDR < 0.1$ , was 278 genes was used instead of the initial 305 genes.

#### [Gene Ontology annotation with DE list of \$FDR < 0.05\$](#)

GO Enrichment Analysis was first performed using the default settings of the function `goana()` on the differentially expressed genes selected, based on an  $FDR < 0.05$  from Section 4.4.3. Results were then limited to GO terms with at least 4 steps back from the root terms for each term, to remove less informative and redundant GO terms. A total of 16 GO terms were statistically significant using FDR-adjusted p-value of less than 0.1. 15 GO terms were associated with the Biological Process and 1 associated with Molecular Function (Figure 5.5). Calcium ion regulation and chemokine response/signalling were the most significant Biological Process terms, whereas cytokine activity was the only significant term from Molecular Function

(Table 5.3). Interestingly, chemokine receptors and ligands such as CXCL10, CCL3 CXCR6, CCR4, CX3CR1 were interconnected between calcium regulation and chemokine responses (Figure 5.6). Studies from Imai et al. (1997) and Al-Aoukaty et al. (1998) have shown that these chemokines are capable of inducing locomotion and mobilization of intracellular calcium and activate the heterotrimeric G proteins, which mediate both leukocyte migration and adhesion. Hence, this suggests a role of ZEB2 in the regulation of calcium homeostasis to support Th1 EM cell migration. Deletion of ZEB2 in Th1 EM cells also results in disruption of the cytokine gene expression including IFNG, CXCL10, BMP4, CCL18, CCL3, TNSFF8 and TNFSF15, similarly described in Section 4.4.3 where IFNG expression was reduced. This is interesting because Th1 EM cells without ZEB2 may not be able to carry out proper effector function owing to potential failure in cytokine secretion and cytokine response, as well as restricted migration to sites of inflammation.

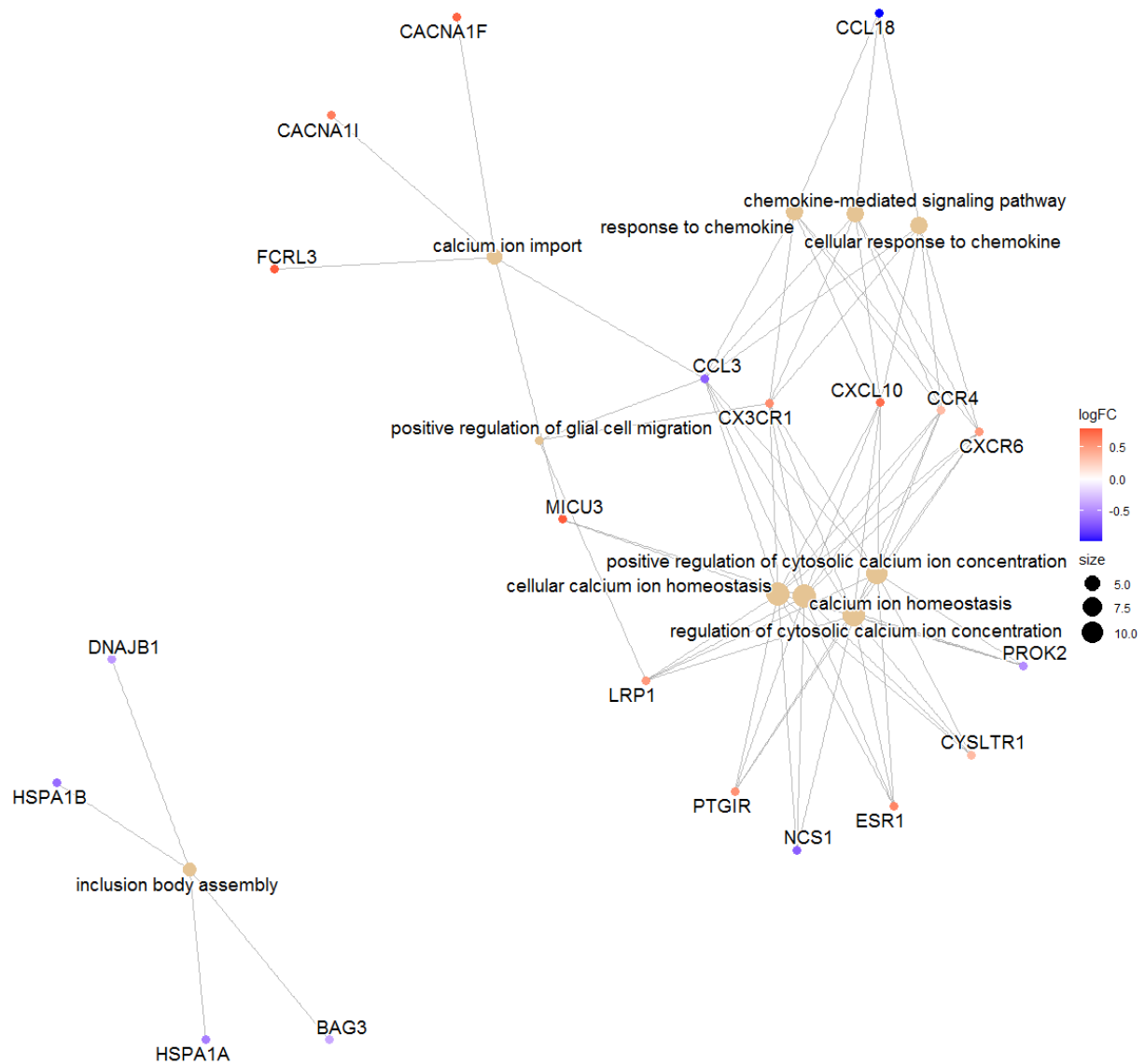


**Figure 5.5: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM pools (Gene list FDR < 0.05)**

Biological Process (BP) and Molecular Function (MF) were annotated as part of the GO terms.

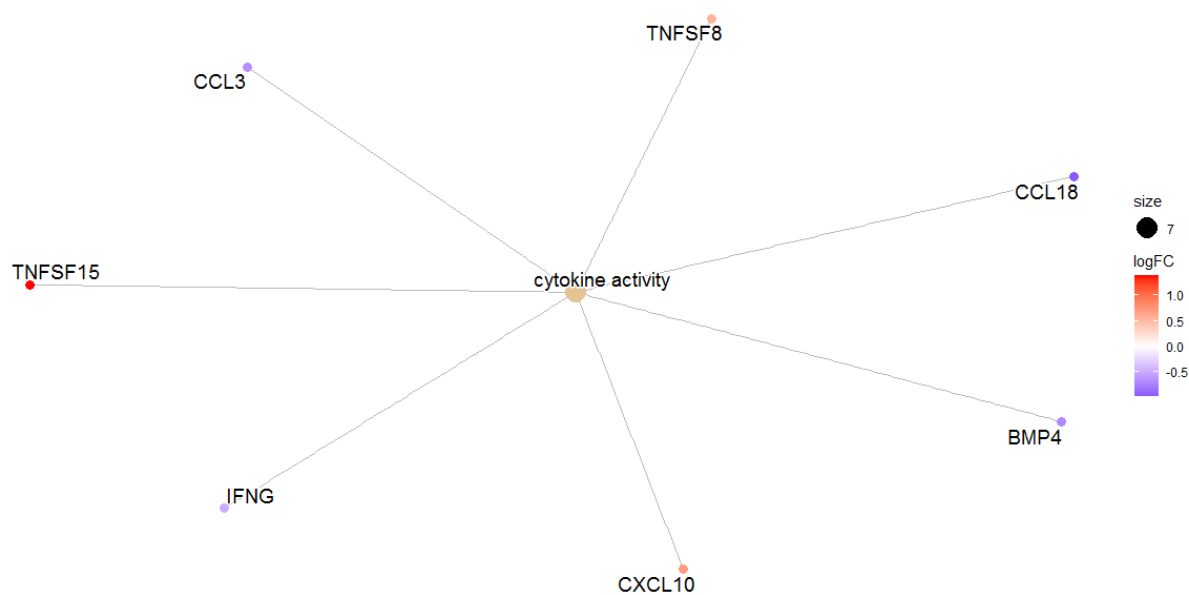
**Table 5.3: Significantly enriched GO terms using an FDR < 0.1 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.05**

GO ID	Term	Ontology	N	Obs DE	Expected DE	FDR
GO:0070098	chemokine-mediated signaling pathway	BP	35	6	0.6	5.40E-02
GO:1990868	response to chemokine	BP	42	6	0.7	5.40E-02
GO:1990869	cellular response to chemokine	BP	42	6	0.7	5.40E-02
GO:0051480	regulation of cytosolic calcium ion concentration	BP	160	11	2.8	5.61E-02
GO:1903977	positive regulation of glial cell migration	BP	8	3	0.1	5.98E-02
GO:0007204	positive regulation of cytosolic calcium ion concentration	BP	150	10	2.6	5.98E-02
GO:0070426	positive regulation of nucleotide-binding oligomerization domain containing signaling pathway	BP	2	2	0	5.98E-02
GO:0070434	positive regulation of nucleotide-binding oligomerization domain containing 2 signaling pathway	BP	2	2	0	5.98E-02
GO:0048104	establishment of body hair or bristle planar orientation	BP	2	2	0	5.98E-02
GO:0048105	establishment of body hair planar orientation	BP	2	2	0	5.98E-02
GO:0070841	inclusion body assembly	BP	21	4	0.4	7.05E-02
GO:0070509	calcium ion import	BP	38	5	0.7	7.20E-02
GO:0042981	regulation of apoptotic process	BP	103 7	33	18.1	7.20E-02
GO:0006874	cellular calcium ion homeostasis	BP	223	12	3.9	7.20E-02
GO:0055074	calcium ion homeostasis	BP	228	12	4	8.22E-02
GO:0005125	cytokine activity	MF	76	7	1.3	6.35E-02



**Figure 5.6: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM pools (Gene list FDR < 0.05)**

CNET plot showing top 10 significantly enriched Biological Process GO terms (orange hub modules) using an FDR < 0.1 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.05.



**Figure 5.7: Gene-concept network plot showing all Molecular Function GO terms**

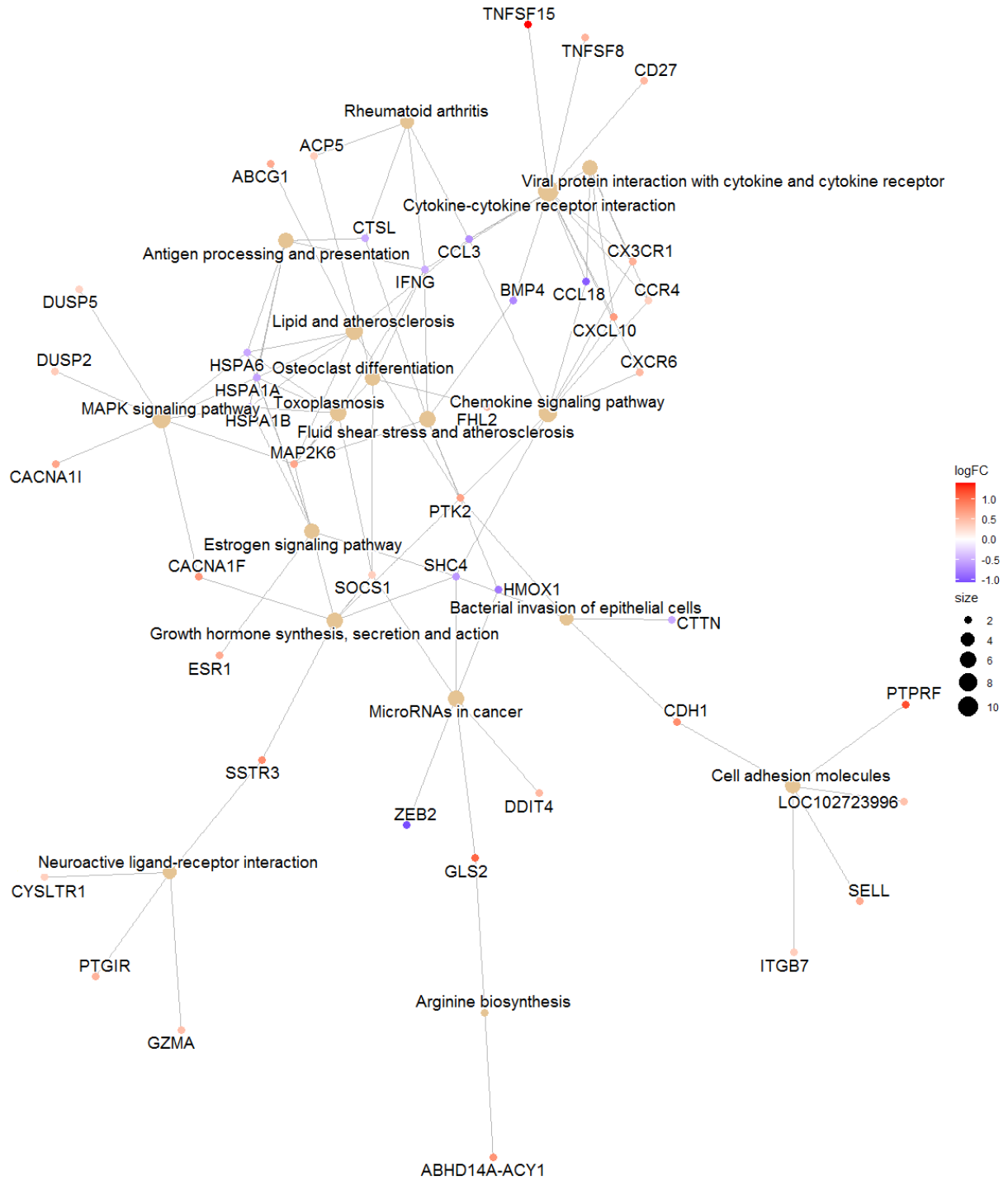
CNET plot showing all significantly enriched Molecular Function GO terms (orange hub modules) using an FDR < 0.1 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.05.

[KEGG pathway annotation with DE list of FDR < 0.05](#)

In order to understand molecular pathways and disease signatures, KEGG pathway analysis was carried out. KEGG pathway analysis carried out on genes differentially expressed between the ZEB2 KO and WT Th1 EM pooled cells resulted in 17 significantly enriched terms. These included “cytokine-cytokine receptor interaction”, “chemokine signaling pathways” and “cell adhesion molecules” suggesting that the ability of Th1 cells to home and migrate is greatly affected by deletion of ZEB2 (Table 5.4). Interestingly, the KEGG term “MicroRNAs in cancer” was also enriched, further validating our enrichment approach since ZEB2 is known to be post-transcriptionally regulated by microRNAs such as miR-155 and miR-200 family in cancer (Park et al., 2008, Bracken et al., 2015, Diaz-Riascos et al., 2019, De Coninck et al., 2019, Brown et al., 2018). Network visualisation of affected genes involved in these significant KEGG terms was as shown in Figure 5.8.

**Table 5.4 Significantly enriched KEGG terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.05**

KEGG ID	Pathway	N	Obs DE	Expected DE	P-value
hsa04060	Cytokine-cytokine receptor interaction	137	11	3.3	2.76E-05
hsa04061	Viral protein interaction with cytokine and cytokine receptor	45	5	1.1	1.08E-03
hsa04062	Chemokine signaling pathway	129	8	3.1	1.91E-03
hsa04612	Antigen processing and presentation	55	5	1.3	2.66E-03
hsa04935	Growth hormone synthesis, secretion and action	85	6	2	3.69E-03
hsa05145	Toxoplasmosis	89	6	2.1	4.63E-03
hsa05418	Fluid shear stress and atherosclerosis	101	6	2.4	8.51E-03
hsa04514	Cell adhesion molecules	78	5	1.9	1.17E-02
hsa04915	Estrogen signaling pathway	83	5	2	1.50E-02
hsa04080	Neuroactive ligand-receptor interaction	61	4	1.5	2.10E-02
hsa04380	Osteoclast differentiation	92	5	2.2	2.18E-02
hsa00220	Arginine biosynthesis	14	2	0.3	2.25E-02
hsa05323	Rheumatoid arthritis	63	4	1.5	2.41E-02
hsa05206	MicroRNAs in cancer	129	6	3.1	2.42E-02
hsa04010	MAPK signaling pathway	204	8	4.9	2.57E-02
hsa05100	Bacterial invasion of epithelial cells	66	4	1.6	2.69E-02
hsa05134	Legionellosis	41	3	1	2.82E-02



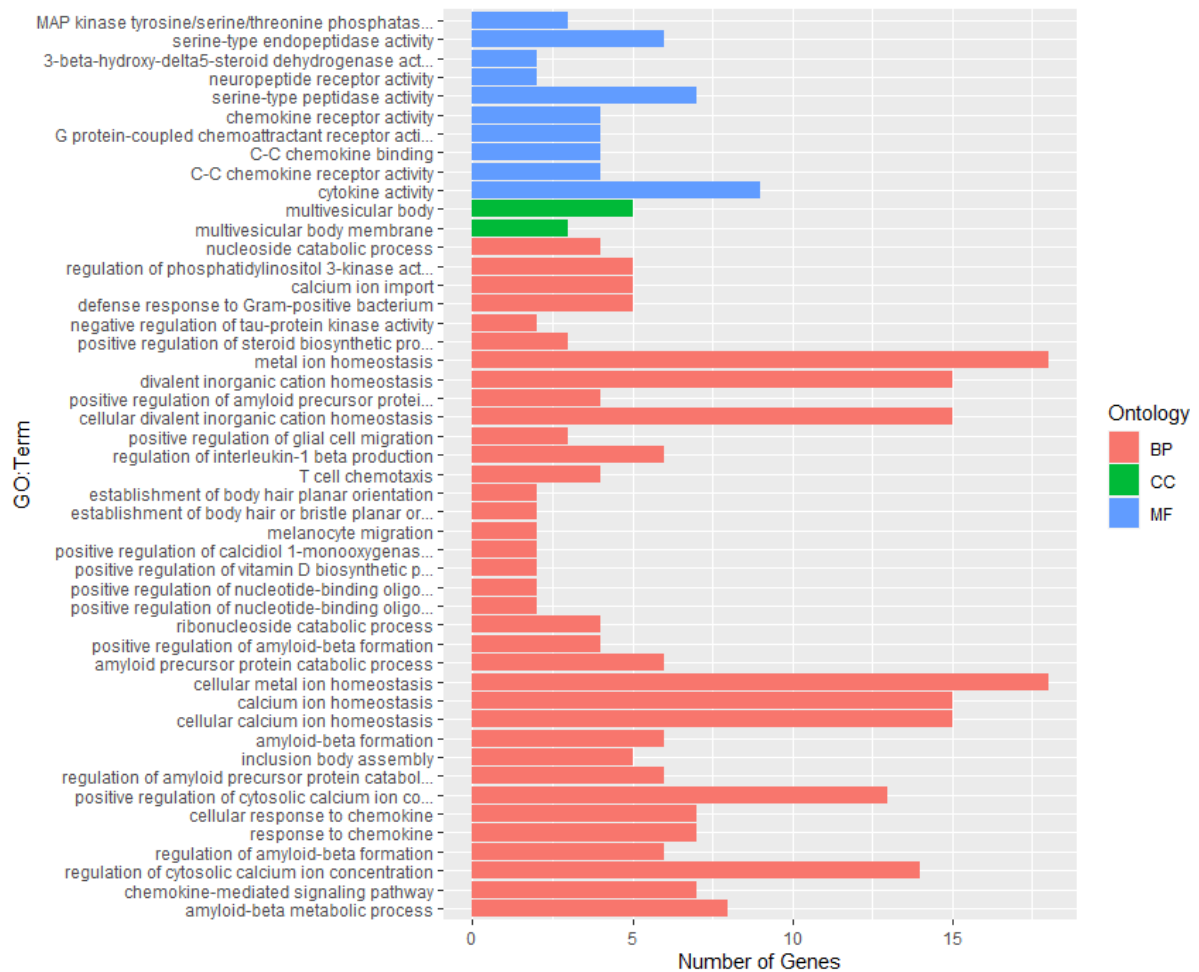
**Figure 5.8: Gene-concept network plot showing all significantly enriched KEGG pathway terms from Th1 EM pools (Gene list FDR < 0.05)**

CNET plot showing all significantly enriched KEGG pathway terms (grey hub modules) using a p-value < 0.05 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.05.



### Gene Ontology annotation with DE list of FDR < 0.1

GO Enrichment Analysis was then performed, again using the DE gene list but with the FDR value loosened to less than 0.1, as mentioned above. There were a total of 44 GO terms that were statistically enriched using the FDR-adjusted p-value of less than 0.1. The numbers of terms from each GO terms were: 10 for Molecular Function, 2 for Cellular Components and 36 for Biological Process (Figure 5.9). Three of the major Biological Process GO term were related to amyloid-beta processing & formation, calcium ion regulation & homeostasis and chemokine response & signalling. This analysis showed similar results to the previous GO term annotation of the gene list using FDR < 0.05 showing the interconnection between calcium regulation and chemokine response as still enriched. Interestingly, the Cellular Component GO term, “multivesicular body” was enriched and one of the downregulated genes, CTSL, is located in the lysosome and is important for IFN $\gamma$  production and Th1 induction in human CD4<sup>+</sup> T cells (Freeley et al., 2018). Lastly, the Molecular Function term, showed enrichment of chemokine activity suggesting that the loss of ZEB2 has affected the ability of the Th1 cells to respond and engage with chemokine activity in order to migrate to sites of inflammation. Affected genes involved in these significant GO terms can be visualized in Figure 5.10 for Biological Process, Figure 5.11 for Molecular Function and Figure 5.12 for Cellular Component.



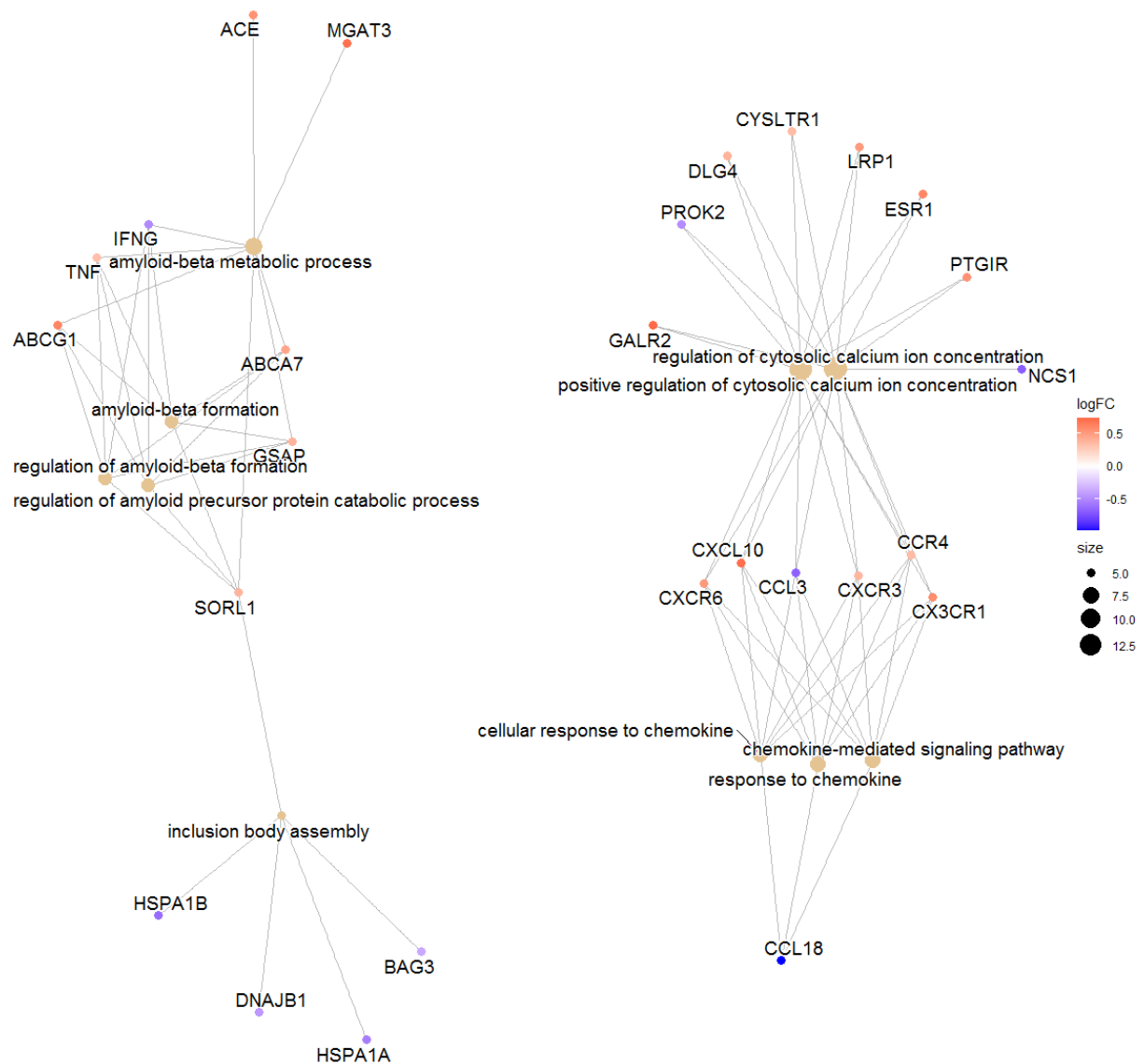
**Figure 5.9: Bar plot showing significant enriched GO terms from ZEB2 KO Th1 EM pools (Gene list FDR < 0.1)**

Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) were annotated as part of the GO terms.

**Table 5.5: Significantly enriched GO terms using an FDR < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1**

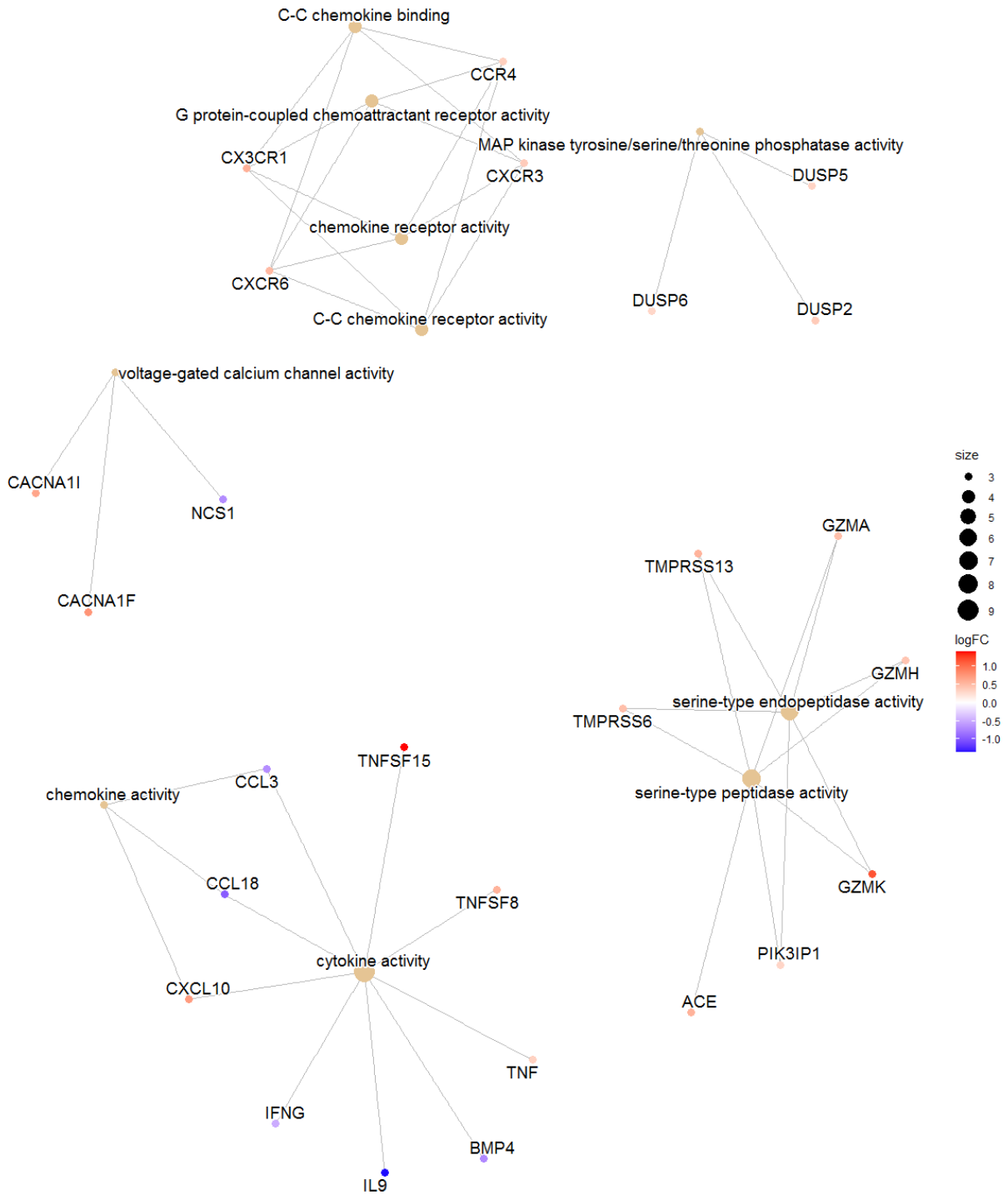
GO ID	Term	Ontology	N	Obs DE	Expected DE	FDR
GO:0050435	amyloid-beta metabolic process	BP	39	8	0.9	7.06E-03
GO:0070098	chemokine-mediated signaling pathway	BP	35	7	0.8	1.79E-02
GO:0051480	regulation of cytosolic calcium ion concentration	BP	160	14	3.8	1.96E-02
GO:1902003	regulation of amyloid-beta formation	BP	27	6	0.6	1.96E-02
GO:1990868	response to chemokine	BP	42	7	1	1.96E-02
GO:1990869	cellular response to chemokine	BP	42	7	1	1.96E-02
GO:0007204	positive regulation of cytosolic calcium ion concentration	BP	150	13	3.6	1.96E-02
GO:1902991	regulation of amyloid precursor protein catabolic process	BP	32	6	0.8	2.38E-02
GO:0070841	inclusion body assembly	BP	21	5	0.5	2.38E-02
GO:0034205	amyloid-beta formation	BP	33	6	0.8	2.38E-02
GO:0006874	cellular calcium ion homeostasis	BP	223	15	5.3	4.79E-02
GO:0055074	calcium ion homeostasis	BP	228	15	5.4	4.79E-02
GO:0006875	cellular metal ion homeostasis	BP	308	18	7.3	4.79E-02
GO:0042987	amyloid precursor protein catabolic process	BP	42	6	1	4.79E-02
GO:1902004	positive regulation of amyloid-beta formation	BP	16	4	0.4	4.79E-02
GO:0042454	ribonucleoside catabolic process	BP	16	4	0.4	4.79E-02
GO:0070426	positive regulation of nucleotide-binding oligomerization domain containing signaling pathway	BP	2	2	0	4.79E-02
GO:0070434	positive regulation of nucleotide-binding oligomerization domain containing 2 signaling pathway	BP	2	2	0	4.79E-02
GO:0060557	positive regulation of vitamin D biosynthetic process	BP	2	2	0	4.79E-02
GO:0060559	positive regulation of calcidiol 1-monooxygenase activity	BP	2	2	0	4.79E-02
GO:0097324	melanocyte migration	BP	2	2	0	4.79E-02
GO:0048104	establishment of body hair or bristle planar orientation	BP	2	2	0	4.79E-02
GO:0048105	establishment of body hair planar orientation	BP	2	2	0	4.79E-02
GO:0010818	T cell chemotaxis	BP	17	4	0.4	4.79E-02
GO:0032651	regulation of interleukin-1 beta production	BP	45	6	1.1	5.06E-02
GO:1903977	positive regulation of glial cell migration	BP	8	3	0.2	5.15E-02
GO:0072503	cellular divalent inorganic cation homeostasis	BP	243	15	5.8	5.15E-02
GO:1902993	positive regulation of amyloid precursor protein catabolic process	BP	19	4	0.5	6.41E-02
GO:0072507	divalent inorganic cation homeostasis	BP	252	15	6	6.83E-02
GO:0055065	metal ion homeostasis	BP	342	18	8.1	8.86E-02
GO:0010893	positive regulation of steroid biosynthetic process	BP	10	3	0.2	8.86E-02
GO:1902948	negative regulation of tau-protein kinase activity	BP	3	2	0.1	9.59E-02
GO:0050830	defense response to Gram-positive bacterium	BP	38	5	0.9	9.62E-02
GO:0070509	calcium ion import	BP	38	5	0.9	9.62E-02
GO:0043551	regulation of phosphatidylinositol 3-kinase activity	BP	38	5	0.9	9.62E-02
GO:0009164	nucleoside catabolic process	BP	23	4	0.5	9.62E-02
GO:0032585	multivesicular body membrane	CC	10	3	0.2	8.86E-02
GO:0005771	multivesicular body	CC	38	5	0.9	9.62E-02
GO:0005125	cytokine activity	MF	76	9	1.8	2.22E-02

GO:0016493	C-C chemokine receptor activity	MF	16	4	0.4	4.79E-02
GO:0019957	C-C chemokine binding	MF	16	4	0.4	4.79E-02
GO:0001637	G protein-coupled chemoattractant receptor activity	MF	17	4	0.4	4.79E-02
GO:0004950	chemokine receptor activity	MF	17	4	0.4	4.79E-02
GO:0008236	serine-type peptidase activity	MF	65	7	1.5	6.12E-02
GO:0008188	neuropeptide receptor activity	MF	3	2	0.1	9.59E-02
GO:0003854	3-beta-hydroxy-delta5-steroid dehydrogenase activity	MF	3	2	0.1	9.59E-02
GO:0004252	serine-type endopeptidase activity	MF	54	6	1.3	9.59E-02
GO:0017017	MAP kinase tyrosine/serine/threonine phosphatase activity	MF	11	3	0.3	9.62E-02



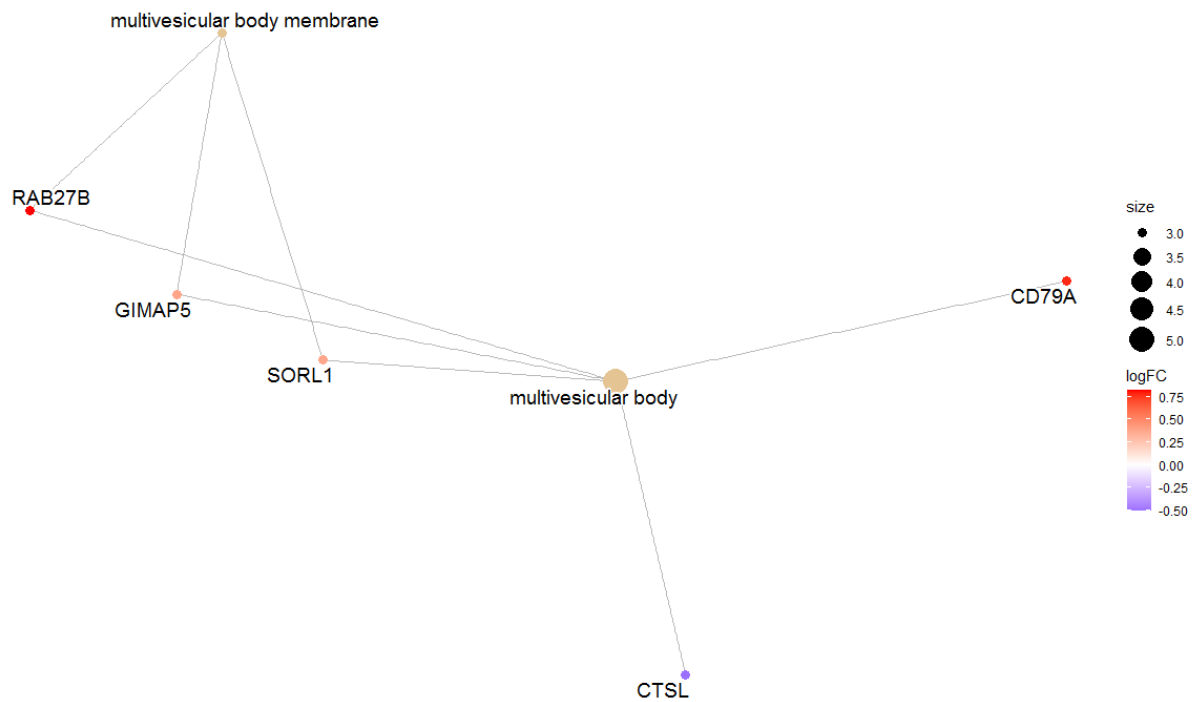
**Figure 5.10: Gene-concept network plot showing top 10 Biological Process GO terms from Th1 EM pools (Gene list FDR < 0.1)**

CNET plot showing top 10 significantly enriched Biological Process GO terms (orange hub modules) using an FDR < 0.1 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.1.



**Figure 5.11: Gene-concept network plot showing all Molecular Function GO terms from Th1 EM pools (Gene list FDR < 0.1)**

CNET plot showing all significantly enriched Molecular Function GO terms (orange hub modules) using an FDR < 0.1 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.1.



**Figure 5.12: Gene-concept network plot showing all Cellular Component GO terms from Th1 EM pools (Gene list FDR < 0.1)**

CNET plot showing top 10 significantly enriched Biological Process GO terms (orange hub modules) using an FDR < 0.1 as the criteria for significance and minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.1.

[KEGG pathway annotation with DE list of FDR < 0.1](#)

KEGG Enrichment Analysis were then performed, as before, using the DE gene list but with the FDR value loosened to less than 0.1 as mentioned earlier. KEGG pathway enrichment analysis showed that, among all the 30 enriched KEGG pathways based on p-value < 0.05, cytokine-cytokine receptor interaction, antigen processing and presentation and autoimmune diseases including Rheumatoid arthritis (RA), Inflammatory bowel disease (IBD) and Type I diabetes mellitus (T1D) were significantly enriched (Table 5.6). Affected genes that are common in those autoimmune disease were HLA-DQA2 and IFNG as shown in Figure 5.13. This analysis confirmed our results and may show evidence that ZEB2 does have a role in autoimmune diseases via dysregulation of multiple genes that are critical in disease pathogenesis when ZEB2 is deleted. Interestingly, there were other KEGG term such as intestinal immune network for IgA production, IL-17 signalling pathways and TGF-beta signalling pathways which further support the link between ZEB2 and IBD.



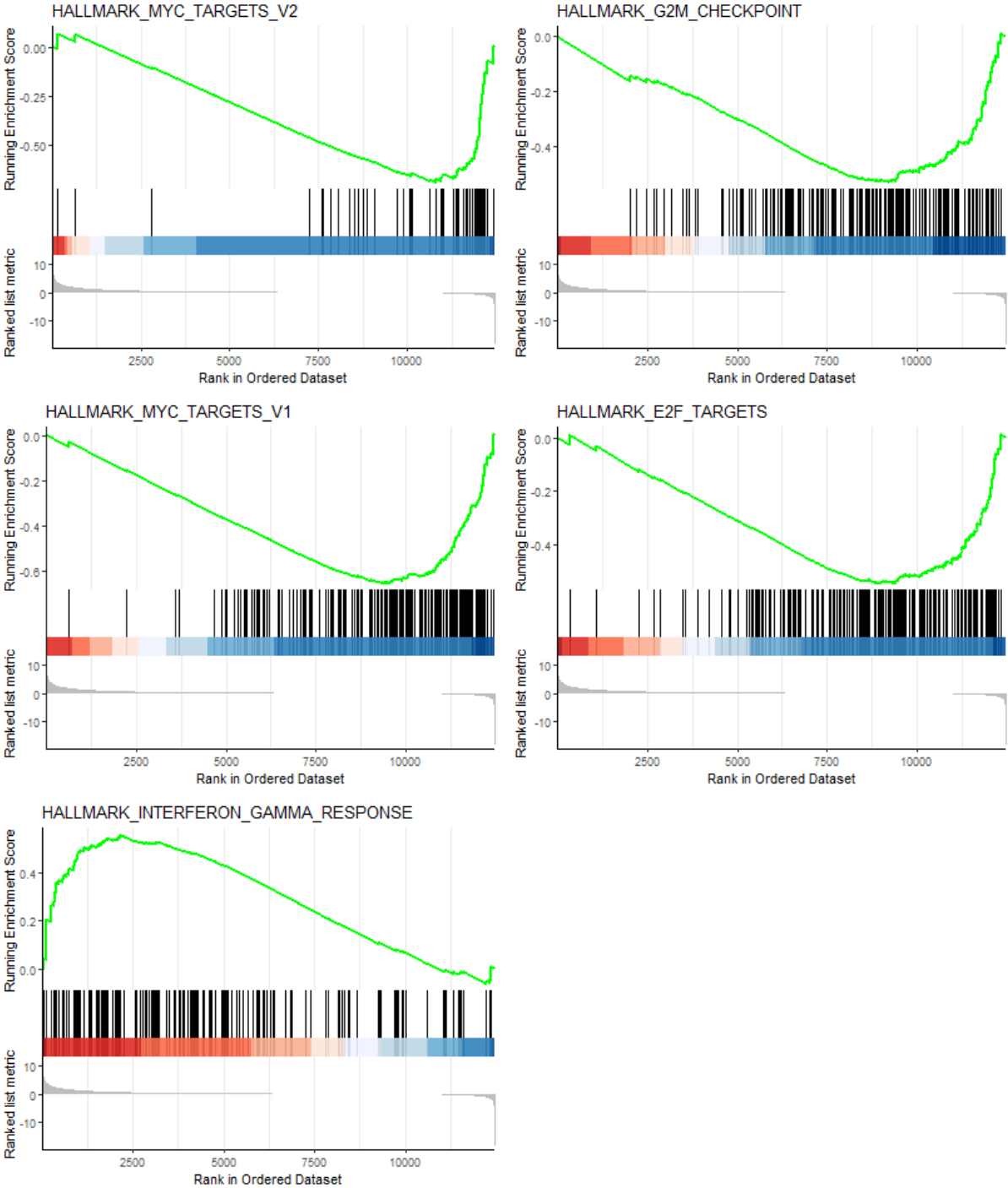
**Table 5.6: Significantly enriched KEGG terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1**

KEGG ID	Pathway	N	Obs DE	Expected DE	P-value
hsa04060	Cytokine-cytokine receptor interaction	137	16	3.2	1.46E-07
hsa04061	Viral protein interaction with cytokine and cytokine receptor	45	8	1.1	9.08E-06
hsa04612	Antigen processing and presentation	55	8	1.3	4.18E-05
hsa05145	Toxoplasmosis	89	9	2.1	2.52E-04
hsa05323	Rheumatoid arthritis	63	7	1.5	6.94E-04
hsa05310	Asthma	19	4	0.5	9.05E-04
hsa04380	Osteoclast differentiation	92	8	2.2	1.50E-03
hsa04514	Cell adhesion molecules	78	7	1.8	2.45E-03
hsa04080	Neuroactive ligand-receptor interaction	61	6	1.4	3.13E-03
hsa04062	Chemokine signaling pathway	129	9	3.1	3.57E-03
hsa05332	Graft-versus-host disease	29	4	0.7	4.60E-03
hsa05321	Inflammatory bowel disease	49	5	1.2	5.88E-03
hsa05330	Allograft rejection	32	4	0.8	6.59E-03
hsa04940	Type I diabetes mellitus	34	4	0.8	8.19E-03
hsa04010	MAPK signaling pathway	204	11	4.8	9.40E-03
hsa04672	Intestinal immune network for IgA production	36	4	0.9	1.00E-02
hsa05418	Fluid shear stress and atherosclerosis	101	7	2.4	1.01E-02
hsa04350	TGF-beta signaling pathway	59	5	1.4	1.28E-02
hsa05144	Malaria	21	3	0.5	1.28E-02
hsa04935	Growth hormone synthesis, secretion and action	85	6	2	1.54E-02
hsa05134	Legionellosis	41	4	1	1.57E-02
hsa05410	Hypertrophic cardiomyopathy	42	4	1	1.71E-02
hsa04390	Hippo signaling pathway	98	6	2.3	2.89E-02
hsa05146	Amoebiasis	50	4	1.2	3.04E-02
hsa05417	Lipid and atherosclerosis	158	8	3.7	3.45E-02
hsa04640	Hematopoietic cell lineage	54	4	1.3	3.88E-02
hsa05142	Chagas disease	79	5	1.9	3.94E-02
hsa00220	Arginine biosynthesis	14	2	0.3	4.22E-02
hsa05140	Leishmaniasis	56	4	1.3	4.34E-02
hsa04657	IL-17 signaling pathway	57	4	1.4	4.59E-02
hsa04915	Estrogen signaling pathway	83	5	2	4.70E-02



### Gene set enrichment analysis with DE list of FDR < 0.1

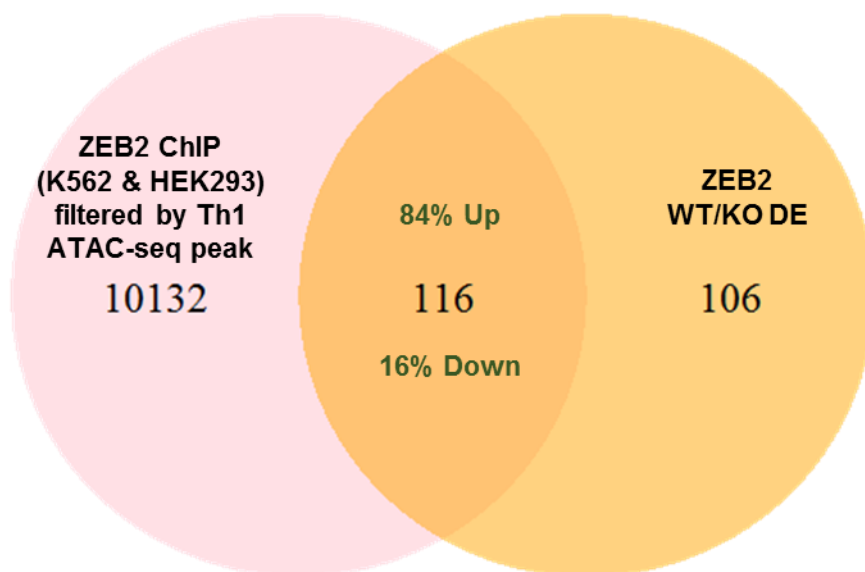
Gene set enrichment analysis (GSEA) is a method to identify classes of genes or proteins that are over-represented in a large set of genes or proteins. For GSEA, I used Hallmark gene sets which summarize and represent specific well-defined biological states or processes and display coherent expression. There were 5 hallmark gene sets that were statistically enriched, and these were: MYC targets, G2M checkpoint, E2F targets and interferon gamma response gene sets (Figure 5.14). Interestingly, even when the expression of IFNG was downregulated in a Th1 EM owing to loss of ZEB2, these cells may potentially have enhanced responses to IFN $\gamma$  signalling leading to strong activation. Loss of ZEB2 showed anti-proliferative signalling (negatively enriched for E2F targets, G2M checkpoint, and Myc targets V2) suggesting the importance of ZEB2 in Th1 cell proliferation and expansion in response to foreign pathogens. This combination of enrichment gene sets is suggestive of a ZEB2 role in regulating Th1 cells in response to IFN $\gamma$  signalling and cell division.



**Figure 5.14: Barcode plot of enriched Hallmark gene sets with GSEA analysis**  
Enrichment plots showing enriched Hallmark gene sets along with running enrichment score that are statistically significant with Bonferroni-adjusted p-value of less than 0.05 for multiple testing.

### 5.4.3 ZEB2 ChIP gene target in Th1

In order to search for putative, direct ZEB2 target genes from my DE genes, I analysed ZEB2 chromatin immunoprecipitation-sequencing (ChIP-seq) datasets from K562 and HEK293 made available by the ENCODE project (There is currently no publicly available T cell ZEB2 ChIP data). Although the ChIP-seq was carried out using non-immune cells, which may not capture genes exclusively expressed in immune cells, this analysis can still infer any DE genes that are potential targets of ZEB2. ZEB2 ChIP-seq peaks were first filtered based on chromatin accessibility from a Th1 ATAC-seq (Calderon et al., 2019). By intersecting filtered ZEB2 ChIP-seq with our ZEB2 KO DE, I found that there were 137 of 222 (~60%) DE genes that are potential direct ZEB2 targets, with 84% genes potentially repressed and 16% genes potentially induced by ZEB2 directly (Figure 5.15). One of the top most significant DE genes in the ZEB2 KO, HMOX1 is also a ZEB2 ChIP target, demonstrating the potential for ZEB2 to interact with those DE genes important for Th1 function (Table 5.7: Direct targets of ZEB2 transcription factor Table 5.7).



**Figure 5.15: Venn diagram showing number of genes that are a target of ZEB2.**

ZEB2 ChIP-seq data merged from both K562 and HEK293 cell lines (Davis et al., 2018) were filtered by chromatin accessibility of an activated Th1 cells (Calderon et al., 2019). The filtered ZEB2 ChIP-seq data (pink) were then annotated to the nearest genes and then intersected with ZEB2 deleted Th1 EM pools DE gene list (FDR < 0.05) (orange) to acquire genes that are a direct target of ZEB2 binding.

**Table 5.7: Direct targets of ZEB2 transcription factor**

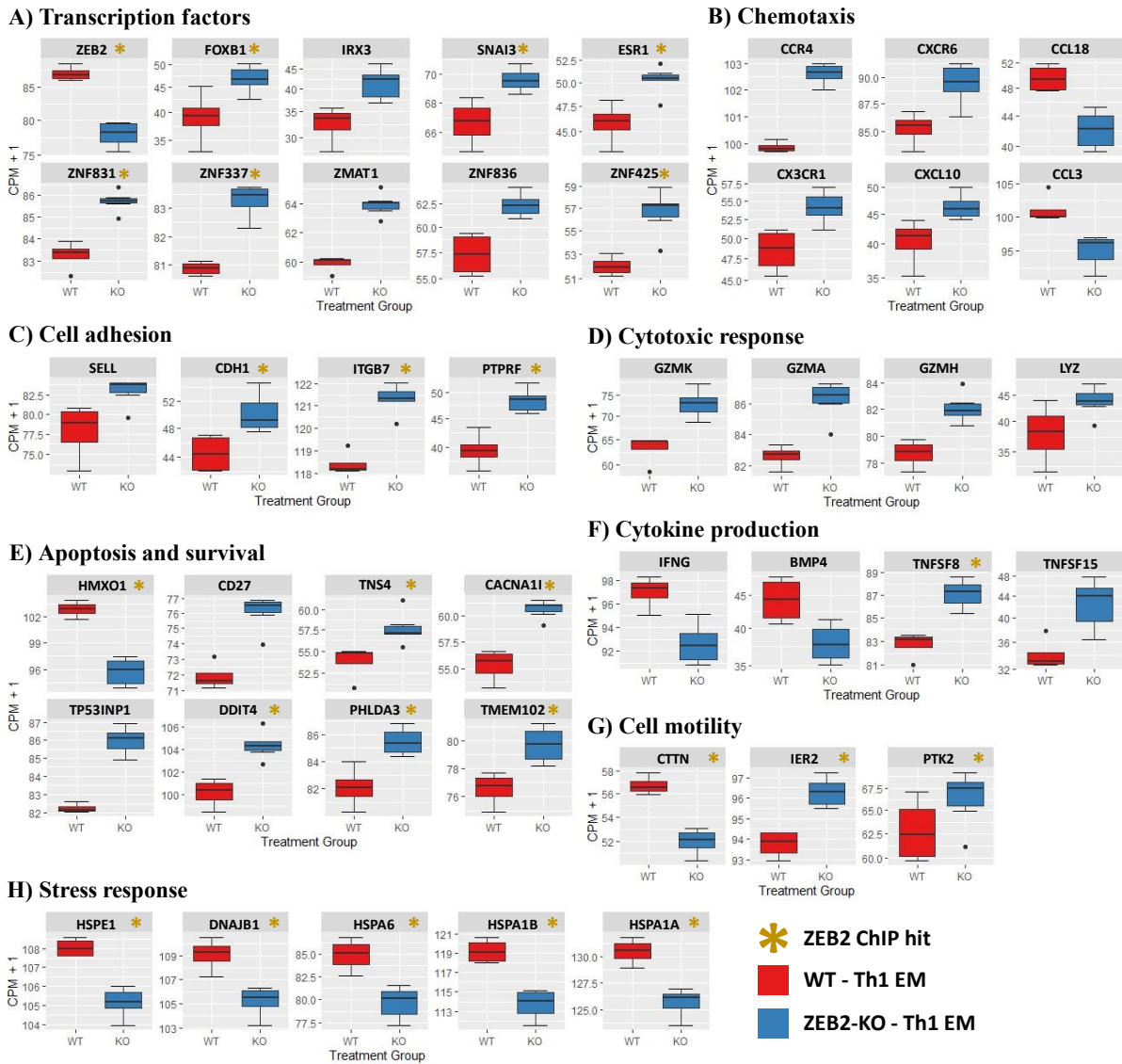
Gene Name	Gene Description	DE Status
ZEB2	zinc finger E-box binding homeobox 2	Down
HMOX1	heme oxygenase 1	Down
TRIM22	tripartite motif containing 22	Up
C12orf57	chromosome 12 open reading frame 57	Down
PTPRF	protein tyrosine phosphatase receptor type F	Up
GTDC1	glycosyltransferase like domain containing 1	Down
GLS2	glutaminase 2	Up
ANKRD24	ankyrin repeat domain 24	Up
PLAC8	placenta associated 8	Up
ABCA7	ATP binding cassette subfamily A member 7	Up
TP53INP1	tumor protein p53 inducible nuclear protein 1	Up
HSPA1B	heat shock protein family A (Hsp70) member 1B	Down
HSPA1A	heat shock protein family A (Hsp70) member 1A	Down
MYO15B	myosin XVb	Up
TNFSF8	TNF superfamily member 8	Up
ABCG1	ATP binding cassette subfamily G member 1	Up
TK2	thymidine kinase 2	Up
CACNA1I	calcium voltage-gated channel subunit alpha1 I	Up
ATP8A2	ATPase phospholipid transporting 8A2	Up
MYOM1	myomesin 1	Up
VWA7	von Willebrand factor A domain containing 7	Up
NABP1	nucleic acid binding protein 1	Up
MAGIX	MAGI family member, X-linked	Up
PLEKHG3	pleckstrin homology and RhoGEF domain containing G3	Up
KCNA2	potassium voltage-gated channel subfamily A member 2	Up
ZDHHC1	zinc finger DHHC-type containing 1	Up
DDIT4	DNA damage inducible transcript 4	Up
CHI3L2	chitinase 3 like 2	Up
C1QTNF6	C1q and TNF related 6	Up
FDXR	ferredoxin reductase	Up
CDH1	cadherin 1	Up
FOXB1	forkhead box B1	Up
CEP126	centrosomal protein 126	Up
PACSIN1	protein kinase C and casein kinase substrate in neurons 1	Up
CDKN2B	cyclin dependent kinase inhibitor 2B	Up
PPP1R15A	protein phosphatase 1 regulatory subunit 15A	Down
CTSL	cathepsin L	Down
AQP3	aquaporin 3 (Gill blood group)	Up
ZNF425	zinc finger protein 425	Up
PLEKHB1	pleckstrin homology domain containing B1	Up
LY9	lymphocyte antigen 9	Up
BAG3	BAG cochaperone 3	Down
CEP68	centrosomal protein 68	Up
CD68	CD68 molecule	Up

ZMAT3	zinc finger matrin-type 3	Up
MAP2K6	mitogen-activated protein kinase kinase 6	Up
DNAJB1	DnaJ heat shock protein family (Hsp40) member B1	Down
PTK2	protein tyrosine kinase 2	Up
GDPD5	glycerophosphodiester phosphodiesterase domain containing 5	Up
FBXO2	F-box protein 2	Up
PTGIR	prostaglandin I2 receptor	Up
ANKRD36C	ankyrin repeat domain 36C	Up
MICU3	mitochondrial calcium uptake family member 3	Up
HSPA6	heat shock protein family A (Hsp70) member 6	Down
RHPN1	rhopilin Rho GTPase binding protein 1	Up
ADAMTS13	ADAM metalloproteinase with thrombospondin type 1 motif 13	Up
TXNIP	thioredoxin interacting protein	Up
NEIL1	nei like DNA glycosylase 1	Up
GAL3ST4	galactose-3-O-sulfotransferase 4	Up
HSPE1	heat shock protein family E (Hsp10) member 1	Down
FHL2	four and a half LIM domains 2	Up
TTC30A	tetratricopeptide repeat domain 30A	Up
FZD6	frizzled class receptor 6	Up
NUDT18	nudix hydrolase 18	Up
VWA5A	von Willebrand factor A domain containing 5A	Up
MYO15A	myosin XVA	Up
PEX7	peroxisomal biogenesis factor 7	Up
ITGB7	integrin subunit beta 7	Up
TREML2	triggering receptor expressed on myeloid cells like 2	Up
APBB3	amyloid beta precursor protein binding family B member 3	Up
SOCS1	suppressor of cytokine signaling 1	Up
ESR1	estrogen receptor 1	Up
FCMR	Fc fragment of IgM receptor	Up
CTTN	cortactin	Down
CARNS1	carnosine synthase 1	Up
ZFAND2A	zinc finger AN1-type containing 2A	Down
PHLDA3	pleckstrin homology like domain family A member 3	Up
ORAI3	ORAI calcium release-activated calcium modulator 3	Up
LRP1	LDL receptor related protein 1	Up
TP53I3	tumor protein p53 inducible protein 3	Up
NPM2	nucleophosmin/nucleoplasmin 2	Up
SLFN5	schlafen family member 5	Up
NCR3LG1	natural killer cell cytotoxicity receptor 3 ligand 1	Down
DPEP2	dipeptidase 2	Up
IER2	immediate early response 2	Up
YPEL3	yippee like 3	Up
B3GALT4	beta-1,3-galactosyltransferase 4	Up
HID1	HID1 domain containing	Down
DUSP2	dual specificity phosphatase 2	Up
WDR60	WD repeat domain 60	Up



TNS4	tensin 4	Up
PLEKHG5	pleckstrin homology and RhoGEF domain containing G5	Up
ZNF879	zinc finger protein 879	Up
PROK2	prokineticin 2	Down
GPR155	G protein-coupled receptor 155	Up
ACP5	acid phosphatase 5, tartrate resistant	Up
ZNF831	zinc finger protein 831	Up
MORN3	MORN repeat containing 3	Up
ARHGEF3	Rho guanine nucleotide exchange factor 3	Up
CDKL2	cyclin dependent kinase like 2	Up
SESN1	sestrin 1	Up
ZNF337	zinc finger protein 337	Up
GHDC	GH3 domain containing	Up
SLCO2B1	solute carrier organic anion transporter family member 2B1	Down
RHOU	ras homolog family member U	Up
BBS2	Bardet-Biedl syndrome 2	Up
CD79A	CD79a molecule	Up
TMEM102	transmembrane protein 102	Up
DUSP5	dual specificity phosphatase 5	Up
TSPAN31	tetraspanin 31	Up
SNAI3	snail family transcriptional repressor 3	Up
SYTL1	synaptotagmin like 1	Up
ADAMTSL5	ADAMTS like 5	Up
VPS9D1	VPS9 domain containing 1	Up
EFEMP2	EGF containing fibulin extracellular matrix protein 2	Up
NICN1	nicolin 1	Up

In order to start to unravel the functional implication of ZEB2 regulated genes, we next extracted significant DE genes from the dataset based on gene families with known physiological roles. Among the DE genes, there were 9 genes that are transcription factors (excluding ZEB2) and 6 of them are ZEB2 ChIP targets. This suggests that ZEB2 serves as a major upstream transcription factor that regulates other downstream transcription factors to control a series of combinatorial transcription networks and hence highlights the importance of ZEB2 in Th1 EM cells. After deletion of ZEB2, these 6 transcription factors were all relieved from ZEB2 repression and upregulated (Figure 5.16A). Chemokine receptors CCR4, CXCR6 and CXCR1 were upregulated, suggesting increased/altered homing potential with the loss of ZEB2, in agreement with my previous enrichment analysis (Figure 5.16B). Interestingly, there was increased expression of cell adhesion molecules including L-selectin (encoded by SELL), E-cadherin (encoded by CDH1) and ITGB7, as well as increased expression of PTPRF, important for regulating focal adhesion (Figure 5.16C). The increased adhesion is supported by decreased Cortactin (encoded by CTTN) important in cell motility (Figure 5.16G). ZEB2 ablation also led to dysregulated production of cytokines such as BMP4, TNSF8 and TNSF15 and increased cytotoxicity by GZMA, GZMK and GZMH. Additionally, genes normally related to cellular stress responses: HSPE1, HSPA6, HSPA1B, HSPA1A and DNAJB1 to allow for better survival, are dysregulated with ZEB2 ablation, resulting in cells more susceptible to cell death and apoptosis (Figure 5.16H & Figure 5.16E). These data suggest that ablation of ZEB2 alters Th1 cells in several fundamentally important ways.



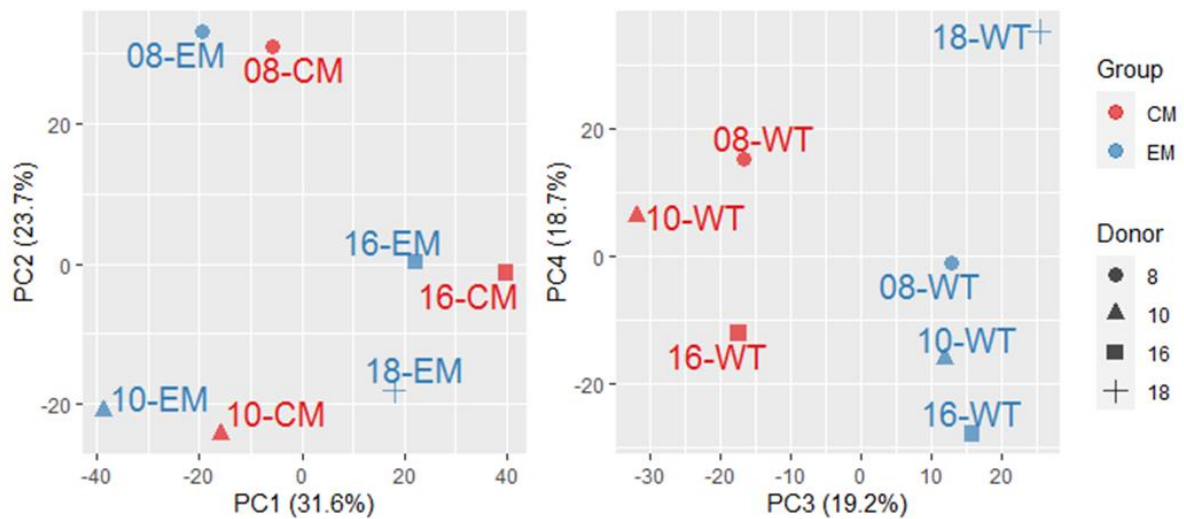
**Figure 5.16: Functional and molecular specialization of DE genes of ZEB2 deleted Th1 EM pools with annotated ZEB2 ChIP**

Expression analysis of transcription factors and genes controlling different functions of Th1 EM cells as indicated. Values represent counts per million + 1 from DE gene list of FDR < 0.05, genes that are ZEB2 ChIP target were annotated with an asterisk.

#### 5.4.4 ZEB2 regulates gene networks to promote Th1 effector memory cell transcriptional programming.

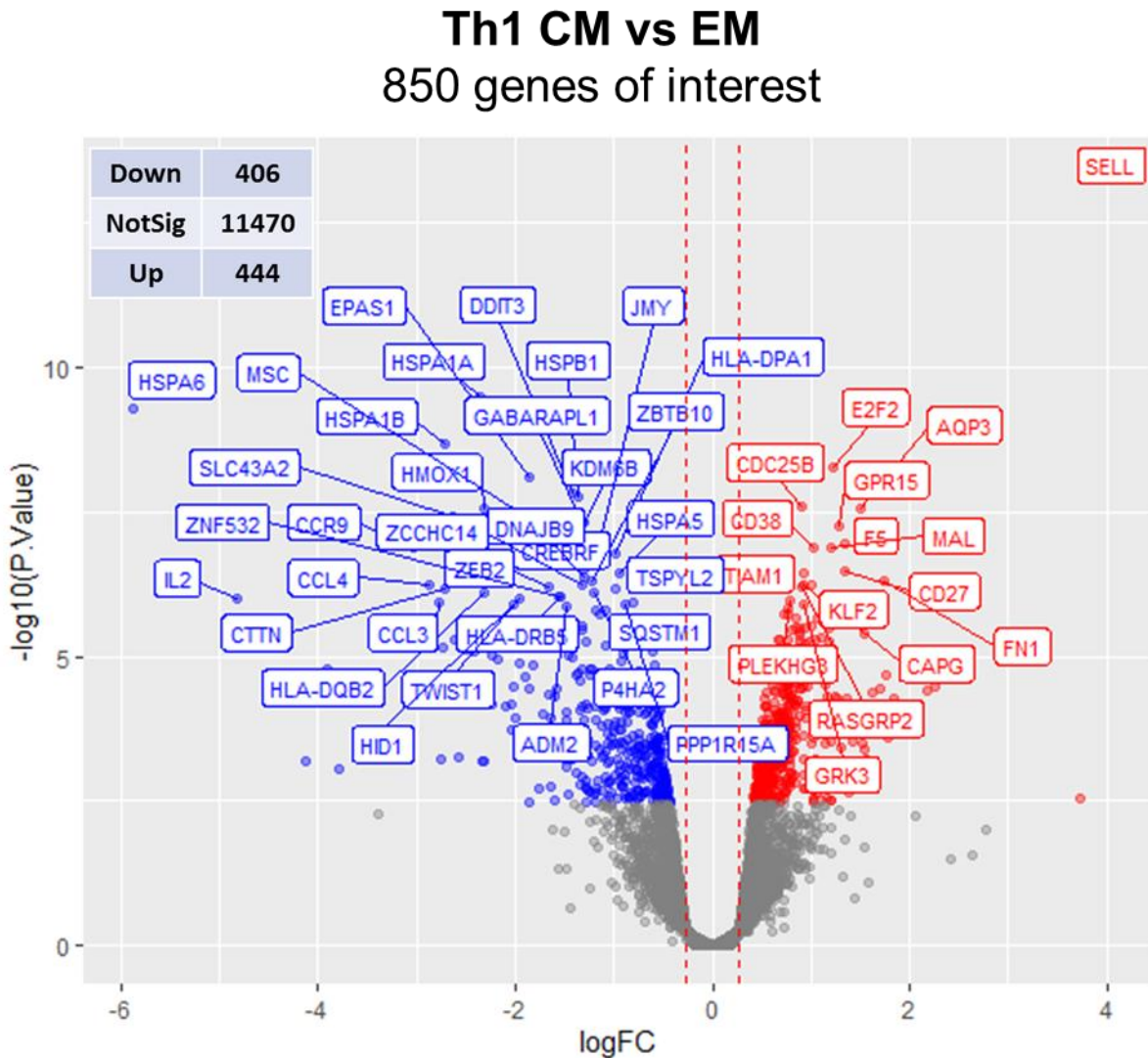
Previous studies by Dominguez et al. (2015) on ZEB2 have shown that ZEB2 is able to repress large numbers of memory signature genes and induce effector signature genes in CD8 T cells. I wished to address this using my data, but first I needed to identify Central Memory (CM) and Effector Memory (EM) signature genes by differential expression analysis comparing WT Th1 EM and WT Th1 CM RNA-seq data from Section 4.4.3. First, I carried out Principle Component Analysis (PCA) on these populations to see how the populations of cells clustered, which gives an idea of how similar or different the populations are to each other. As seen in the left PCA plot from Figure 5.17, 7 samples (3 Th1 CM and 4 Th1 EM) were clustered based on donor on PCA1 and PCA2, suggesting a strong donor variation. When samples were broadcast on PCA3 and PCA4, samples clustered based on memory status (Th1 CM or Th1 EM), indicating that the underlying effect of memory status was still evident but to a lesser extent (Figure 5.17 right). I then defined the Th1 CM and Th1 EM gene expression signature by identifying genes differentially expressed between Th1 CM and Th1 EM cells using an FDR < 0.05 and a cut-off of 1.5 log<sub>2</sub>-fold change shown in the volcano plot (Figure 5.18). There were 850 genes that met these criteria, of which 444 were up-regulated in Th1 CM relative to Th1 EM cells, whereas conversely, 406 genes were up-regulated in Th1 EM relative to Th1 CM cells (Figure 5.18). Next, I examined the set of genes differentially expressed in common between ZEB2-KO/WT Th1 EM from Section 4.4.3 as the ZEB2-dependent gene set (Figure 5.19). Genes that are in this set depend on the presence of ZEB2 to cause a change in phenotype (CM to EM). These analyses show that within the ZEB2 KO Th1 EM, there was a marked increase in the expression of several CM-signature genes and a decrease in the expression of several EM-signature genes (Figure 5.20). Interestingly, a few Th1 EM signature genes such as CXCR6 and GZMH were upregulated in ZEB2-KO Th1 EM suggesting that the loss of ZEB2

increases those genes that are normally repressed in a Th1 CM such as *SELL* (encoding CD62L) and *CDH1* (encoding E-cadherin) (Figure 5.20). Overall, these data support the notion that ZEB2 regulates a small but significant number of EM and CM genes that are important for Th1 effector memory differentiation.



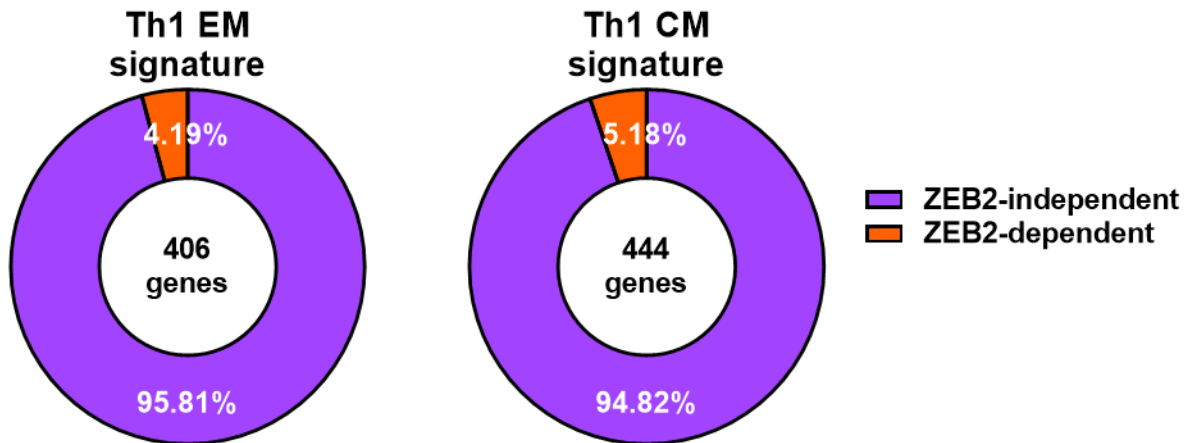
**Figure 5.17: PCA of Th1 CM & Th1 EM cells**

Principal component analysis of all samples run on all expressed genes. PC1-PC2 (left) showed samples clustered by donor and PC3-PC4 (right) showed samples clustered by experimental group.



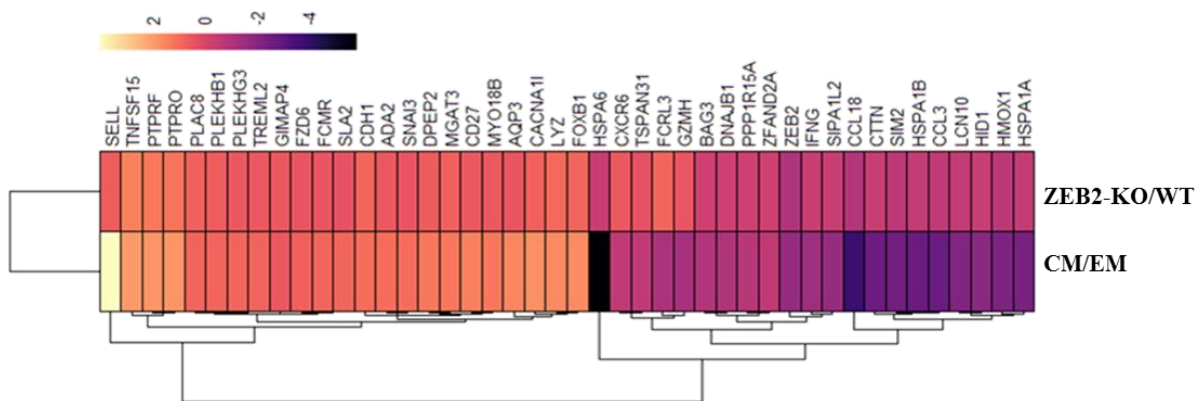
**Figure 5.18: Volcano plot comparing Th1 CM vs Th1 EM cells**

Volcano plot displaying differential expressed genes between Th1 CM and Th1 EM. The vertical axis (y-axis) corresponds to the mean expression value of  $-\log_{10}(\text{PValue})$ , the horizontal axis (x-axis) displays the  $\log_2$  fold change value, the red vertical dotted lines indicate  $\log_2$  fold change of 1.5. The red dots represent the up-regulated expressed transcripts; the blue dots represent the transcripts whose expression downregulated. The top 50 genes that were differentially expressed were labelled.



**Figure 5.19: ZEB2 regulates a subset of EM and CM signature genes**

Pie graphs show the total numbers of EM and CM-signature genes subdivided by the frequency of genes that are dependent on ZEB2 for normal expression.



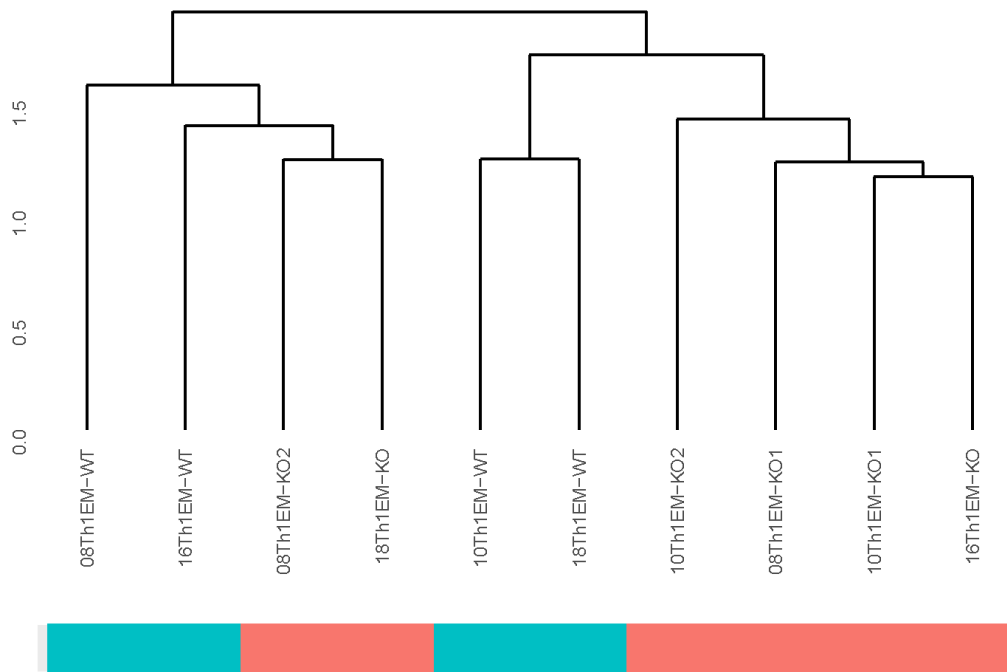
**Figure 5.20: Heatmap of CM and EM signature genes that are regulated by ZEB2**

Hierarchical clustering heatmap analysis of CM and EM signature genes that are regulated by ZEB2 in Th1 EM. Heatmaps are based on log-transformed expression values, are z-scaled by rows and were plotted using the R package pheatmap. All genes and samples were clustered with Euclidean distance.

### 5.4.5 Gene co-expression network analysis with CEMiTool

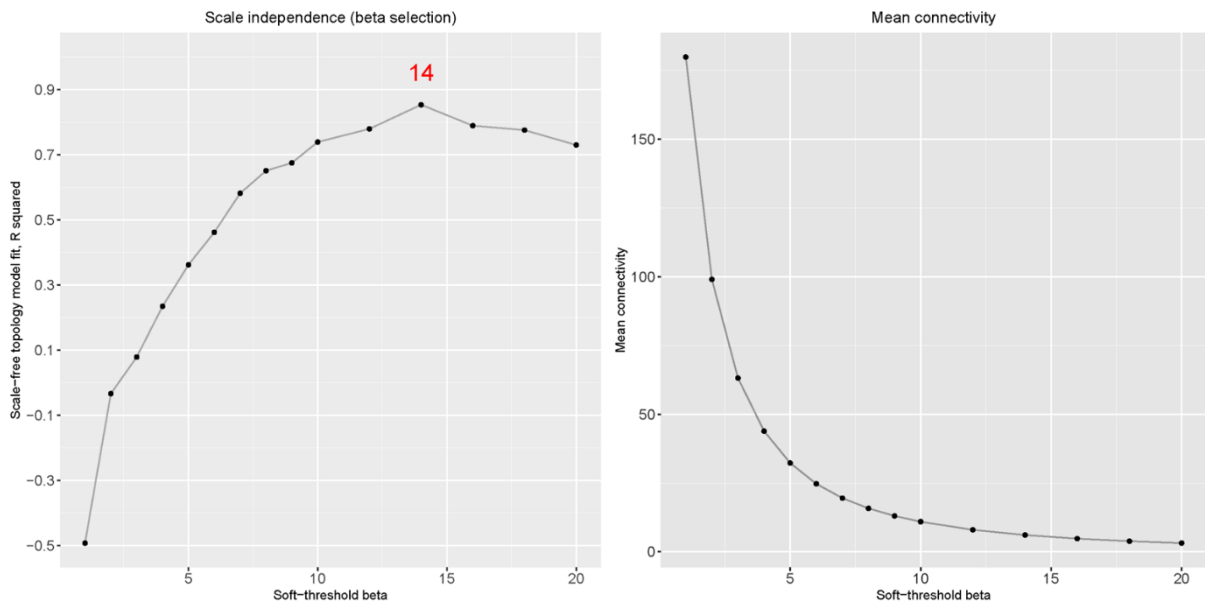
Co-Expression Module identification (CEMi) was then carried out with the aim of investigating modules of genes that are related to the role of ZEB2. It groups genes based on their relatedness and then superimposes in WT and KO data to look for significant difference of the modules. Highly co-expressed gene modules were inferred from all the normalized gene counts from Section 4.4.3. In addition, as I was not interested in stably expressed genes across samples, these genes were excluded from co-expression analysis, leaving 456 genes. Clustering dendrograms of samples based on the expression profiles are shown in Figure 5.21. Even though 4 samples (08Th1EM-KO2, 18Th1EM-KO, 10Th1EM-WT and 18Th1EM-WT) were clustered away from their groupings, these samples were still kept in the analysis in order to increase the power of module detection. A soft-thresholding value is used in CEMiTool to construct modules of co-expressed genes. The soft-thresholding value plays a critical role in determining the number of co-expressed modules, and thus has a significant impact on the findings. In the current analysis, soft thresholding power  $\beta$  value of 14 was used, as it was the lowest power  $\beta$  to make the curve saturate in both  $R^2$  and Mean connectivity plot against soft-threshold  $\beta$  (Figure 5.22).





**Figure 5.21: Sample clustering tree (dendrogram).**

This dendrogram aims to show if there are closely related groups within samples where blue indicates Th1 EM WT group and red indicates Th1 EM KO group.



**Figure 5.22: Beta x R<sup>2</sup> plot and Mean connectivity plot.**

The Beta x R<sup>2</sup> plot is used to visualize the selection of the soft-thresholding parameter  $\beta$  and its corresponding adherence to the scale-free topology model. Selected  $\beta$  values is shown in red. The Mean connectivity plot is intended to show the trade-off between the network’s underlying connectivity and a higher adherence to the scale-free topology model (via higher values of the soft-threshold  $\beta$  parameter).

The module analysis produced 3 significant co-expression modules and allocated the top genes with the highest connectivity in each module (Table 5.8 & Table 5.9). In module 1 (M1), there were 142 genes and the top genes were GTDC1, PLCH2, HID1, GZMK and LCN10 (Table 5.10). In module 2 (M2), there were 127 genes and the top genes were APBA1, CA2, SLC12A7, NME4 and CD200 (Table 5.11). In module 3 (M3), there were 88 genes and the top genes were GSN, MGAT3, STAP2, FAAH2 and MTCO1P12 (Table 5.12). After sample annotation is added, the CEMiTool function automatically evaluates how the modules are up or down regulated between Th1 EM WT and Th1 EM KO. This was performed using the gene set enrichment analysis function from the fgsea package and the enrichment score from Table 5.8. This generated a heatmap for visualization (Figure 5.23). This analysis showed that modules 1 and 3 were significantly more enriched in Th1 EM KO than Th EM WT, whereas module 2 is the opposite. In order to determine which biological functions are associated with the modules, Over Representation Analysis (ORA) was performed with this Molecular Signature Database (MSigDB) hallmark pathway list. Interestingly both M1 and M3 were enriched for inflammatory responses, which further supports our previous findings that Th1 EM KO cells are more inflammatory than their WT counterpart (Figure 5.24 & Figure 5.26). The Th1 cytokine, IFN $\gamma$  was reduced, seemingly contradicting the observation of increased inflammatory response, but this may be owing to a ZEB2 KO specific increase of other inflammatory responses that are not classically Th1 responses. In addition, genes in the M2 module may associate with TNF- $\alpha$  signalling via NF- $\kappa$ B and this module was found to be less enriched in the Th1 EM KO compared with Th1 EM WT (Figure 5.25). As diminished NF- $\kappa$ B was shown to impair Th1 responses and abrogate IFN $\gamma$  production, therefore this further supports our finding that the loss of ZEB2 in Th1 EM may lead to loss of Th1 function (Aronica et al., 1999).

**Table 5.8: Modules enrichment score with adjusted p-value**

Modules	Th1 EM KO: Adjusted p-value	Th1 EM KO: NES	Th1 EM WT: Adjusted p-value	Th1 EM WT: NES
M1	2.42E-08	2.26	1.14E-08	-2.30
M2	9.01E-05	-1.80	5.13E-05	1.84
M3	2.19E-02	1.51	1.92E-02	-1.53

**Table 5.9: Modules information report**

Module	No.Genes	Hubs
M1	142	GTDC1, PLCH2, HID1, GZMK, LCN10
M2	127	APBA1, CA2, SLC12A7, NME4, CD200
M3	88	GSN, MGAT3, STAP2, FAAH2, MTCO1P12

**Table 5.10: Genes in M1 module**

HMOX1	CCL3	CXCR6	C12orf57	BNIP3L	CCL4
GTDC1	ZEB2	FOXP3	ABCG1	PLCH2	CD27
CD68	FAM156B	CTSL	KCNA2	SDR42E1	CAMSAP2
MYO15B	SIRPG	ALDOC	ARSD	FAM3C2	GZMK
PLEKHB1	CBWD6	RHOA	TREML2	AC008878.3	RBM14.RBM4
MSH5.SAPCD1	TSPAN18	SLX1B	PRR29	AC104452.1	IL13
NEIL1	RGPD5	H4C14	CCR6	SPR	USP32P2
RPL17.C18orf32	NCS1	AC116366.3	PLAC8	C3AR1	ZMAT1
KCTD15	AC010323.1	TP53I3	CTNND1	SCNN1D	EPHA4
SPRED2	HLA.DQA2	KLHL3	CDKN2B	EGLN3	LZTS1
TMTC2	TMEM30B	TLR1	ZNF836	GOLGA8N	MYO18B
ZNF626	PCSK1N	NPIPA5	MATK	FP565260.3	ATP6V1G2.DDX39B
CDRT4	AL132656.4	RILP	CEP126	ALG14	CDK5R1
ABCB9	IQCN	PHLDB1	CACNA1I	PTPRN2	ADAMTSL5
SYP	NAV1	ANKRD24	BACE1	FSD1	CBS
AC098484.3	FAM47E.STBD1	AMIGO3	TRO	PFKFB4	TCAF1P1
VWCE	AL031846.1	LRP1	MPZL2	MTCP1	L3HYPDH
AC008764.4	CAPN12	RAB19	GPR161	ZNF658	RASGEF1B
FSBP	TNS4	CADM1	USP6NL	CALD1	BBS5
B3GNT10	TENT5A	PLEKHG5	EHHADH	SALL2	TRIM73
LENG9	ZNF879	P2RY14	ZNF443	GPR75	HID1
LCN10	ZNF563	TMEM8B	CFP	ADAM22	EBI3
TMEM121	CYSLTR2	RPS10P7	ZNF425	C4orf50	AKAP3
PTGIR	MISP3	TJP3	NAIPP4		

**Table 5.11: Genes in M2 modules**

SES3	TCTN3	SLC12A7	HLA.DQB1	MATR3.1	GJB2
APBA1	AL669918.1	ADA2	IL9R	PWP2	DERPC
GNLY	PLXND1	APOL4	CTSW	ADTRP	NME4
TRAV9.2	TRBV5.1	AC010132.3	MTND2P28	TRBV7.9	ATP5MF.PTCD1
AC022384.1	AC073896.1	AC005943.1	AL139300.1	AC107871.1	FBXO2
IRS1	AMY2B	CFAP298.TCP10L	GCSH	ARHGAP11B	TUBB2A
KCNC3	FDXACB1	TMED7.TICAM2	AP002990.1	PTK2	ANXA2R
TRIM34	CACNA2D4	UPK3BL2	PPARGC1B	TEN1.CDK3	F8A3
TRGC2	TRAV29DV5	GSTM3	TSPAN33	CASS4	TRBV3.1
AL022238.4	MUC20	TMEM189.UBE2V1	CTSF	IFIT3	ZNF826P
GREM2	RGPD2	DBN1	PHC1P1	GLB1L2	NHSL2
FOSL1	TIGD1	JAKMIP2	CHAC1	CAMKK1	TRGC1
AKR1C3	TMOD1	ICA1L	CD200	GNA11	MT1F
MACROH2A2	NR4A3	SIPA1L2	RWDD2A	KLRG1	RNLS
AMIGO1	UST	LRR37A	LINC00672	BCL2A1	ZNF677
CHRNA5	FGL2	C2orf76	AC007191.1	SERPINE2	TRAV27
IL21	PLEKHA7	SEN3.EIF4A1	CMTM4	HSF5	MFSD2A
ANPEP	ELOVL4	GABRR2	MINAR1	FGFBP2	ADAMTS10
CMPK2	CA2	TBC1D32	CHN1	C7orf25	LONRF3
AC242376.2	PAQR6	PLAUR	RASGEF1A	ATG9B	CD9
AC009412.1	HLA.DOB	RNASEK.C17orf49	MYRF	CDON	CACNA2D2
CDC42BPA					

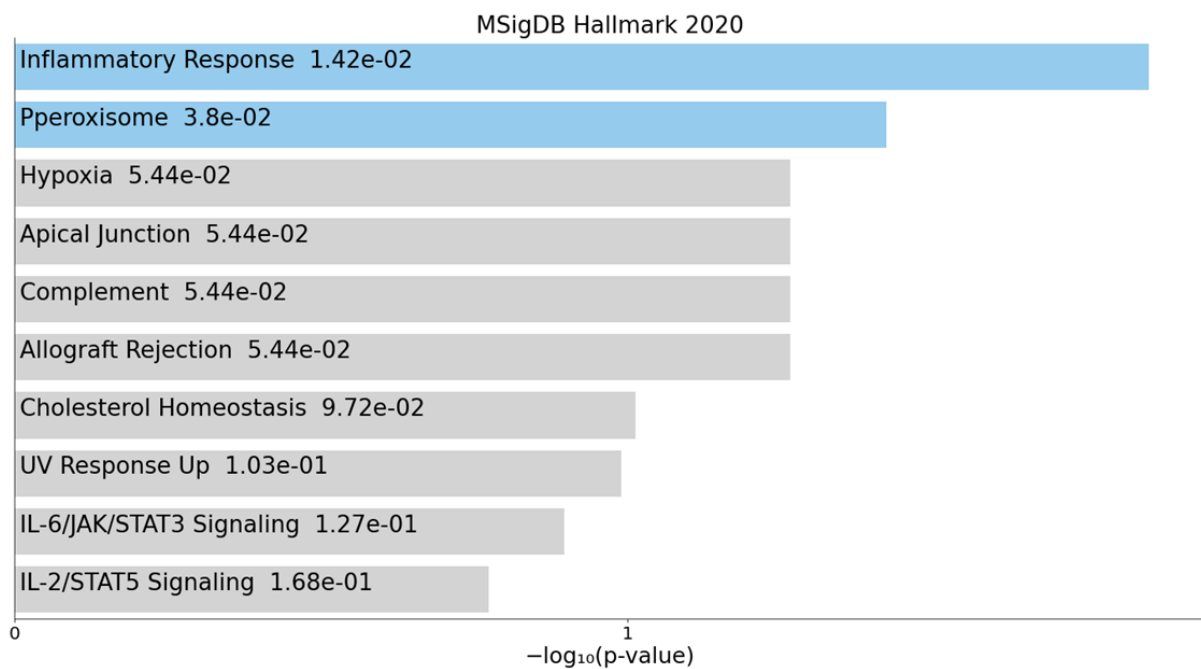
**Table 5.12: Genes in M3 modules**

ARRDC3	HSPA6	AK4	SPINT2	DNM1P47	HBEGF
NAPSA	EML5	IL1R2	AL121594.1	H3P6	HSF4
AL662899.4	FBLN7	C20orf204	MTCO1P12	PLXNB1	EDA2R
LXN	HLA.L	ABCA1	AP003419.1	EIF3FP3	ACCS
LY75.CD302	N4BP3	IFIT2	LRR37A3	SMIM11A	SMAD1
CARNS1	COL18A1	NSG1	MT1G	STAP2	TYW1B
SVIL2P	ID3	ADGRE1	CBY1	CXorf21	TMEM44
TMPRSS6	GSN	TNXB	MGAT3	CITED4	TTN
ZNF815P	STX16.NPEPL1	ZNF629	AL136295.4	VLDLR	IL18RAP
PRR22	TAGLN	KANK3	DNAJC25.GNG10	ZNF620	PEX7
FAAH2	AMOTL1	AL590764.2	ZNF793	IQCC	AC138811.2
GALNT12	NLRP6	CNTNAP1	AC083899.1	ABHD17C	ASTN2
PROK2	LRR37A3	DYNC2H1	NT5E	MORN3	OTOF
MRPL45P2	LAMP3	ARMCX2	MYO5C	GBGT1	GRASP
FN1	PECAM1	AC233968.1	TRAPPC2B		



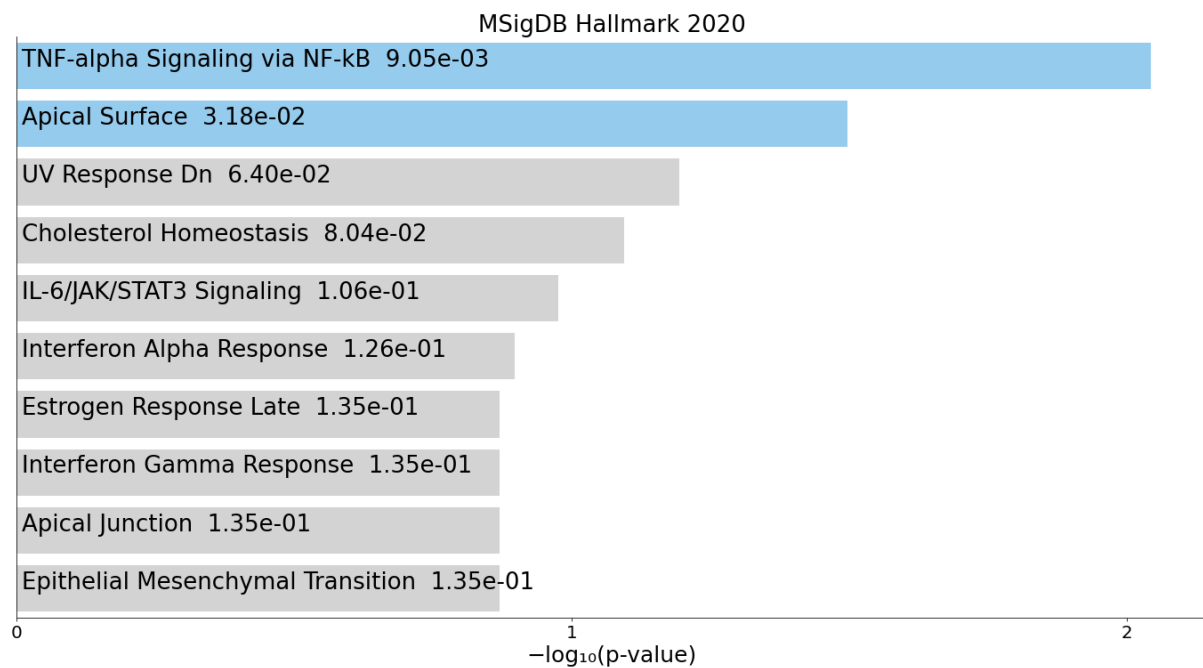
**Figure 5.23: Heatmap of module enrichment analysis**

The enrichment of the modules varies across Th1 EM WT and KO. The size and intensity of the circles in the figure correspond to the Normalised Enrichment Score (NES), which is the enrichment score for a module in each class normalised by the number of genes in the module.



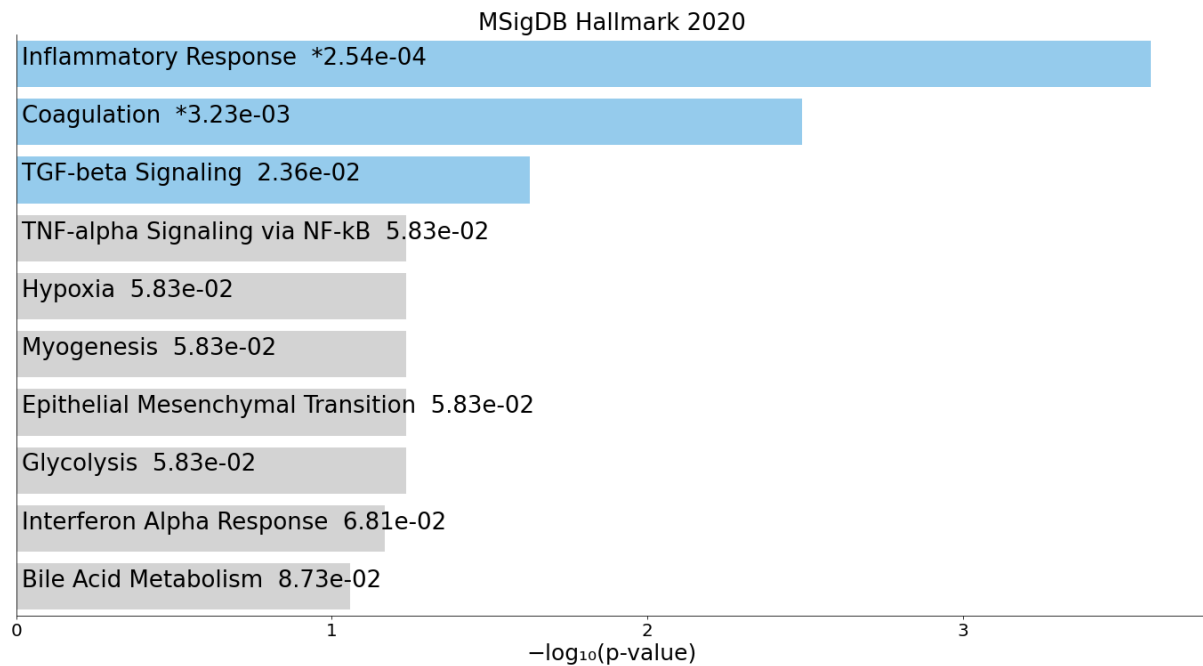
**Figure 5.24: Over Representation Analysis of modules M1 using gene sets from the MsigDB Hallmark Pathway database.**

Bar graph shows the  $-\log_{10}$  adjusted P-value of the enrichment between genes in modules and gene sets from MsigDB Hallmark Pathway database. The pathways were ordered by significance as indicated in the x-axis. Blue bar indicates statistically significant based on adjusted P-value of 0.05.



**Figure 5.25: Over Representation Analysis of modules M2 using gene sets from the MSigDB Hallmark Pathway database.**

Bar graph shows the  $-\log_{10}$  adjusted P-value of the enrichment between genes in modules and gene sets from MSigDB Hallmark Pathway database. The pathways were ordered by significance as indicated in the x-axis. Blue bar indicates statistically significant based on adjusted P-value of 0.05.



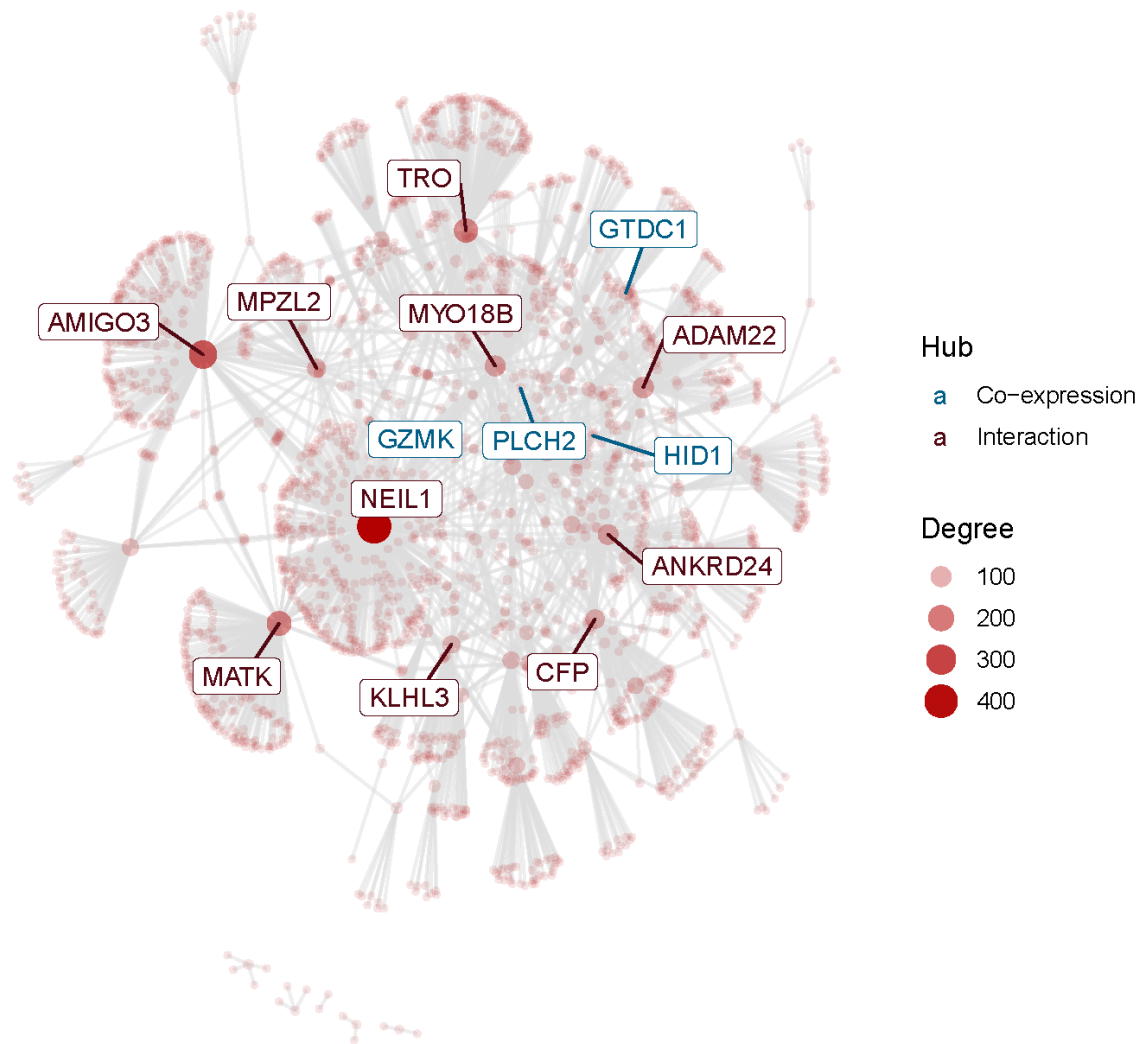
**Figure 5.26: Over Representation Analysis of modules M3 using gene sets from the MSigDB Hallmark Pathway database.**

Bar graph shows the  $-\log_{10}$  adjusted P-value of the enrichment between genes in modules and gene sets from MSigDB Hallmark Pathway database. The pathways were ordered by significance as indicated in the x-axis. Blue bar indicates statistically significant based on adjusted P-value of 0.05.

The CEMiTool also integrates co-expression analysis with protein-protein interaction data. Expression of important genes associated with inflammatory responses that showed interactions, including genes encoding NEIL1, MYO18B and ANKRD24 from M1 module as well as TAGLN, DYNC2H1 and MGAT3 were identified as hubs in module M3 (Figure 5.27 & Figure 5.28). These genes, involved in inflammatory responses, were all upregulated when ZEB2 is lost in Th1 EM and possibly compensate for the loss of the Th1 effector response. Altogether, the co-expression analysis of Th1 EM WT vs. KO cells revealed the 3 modules that were differentially enriched result in increased non-Th1 inflammatory responses owing to a decrease in TNF- $\alpha$  signalling via NF- $\kappa$ B.



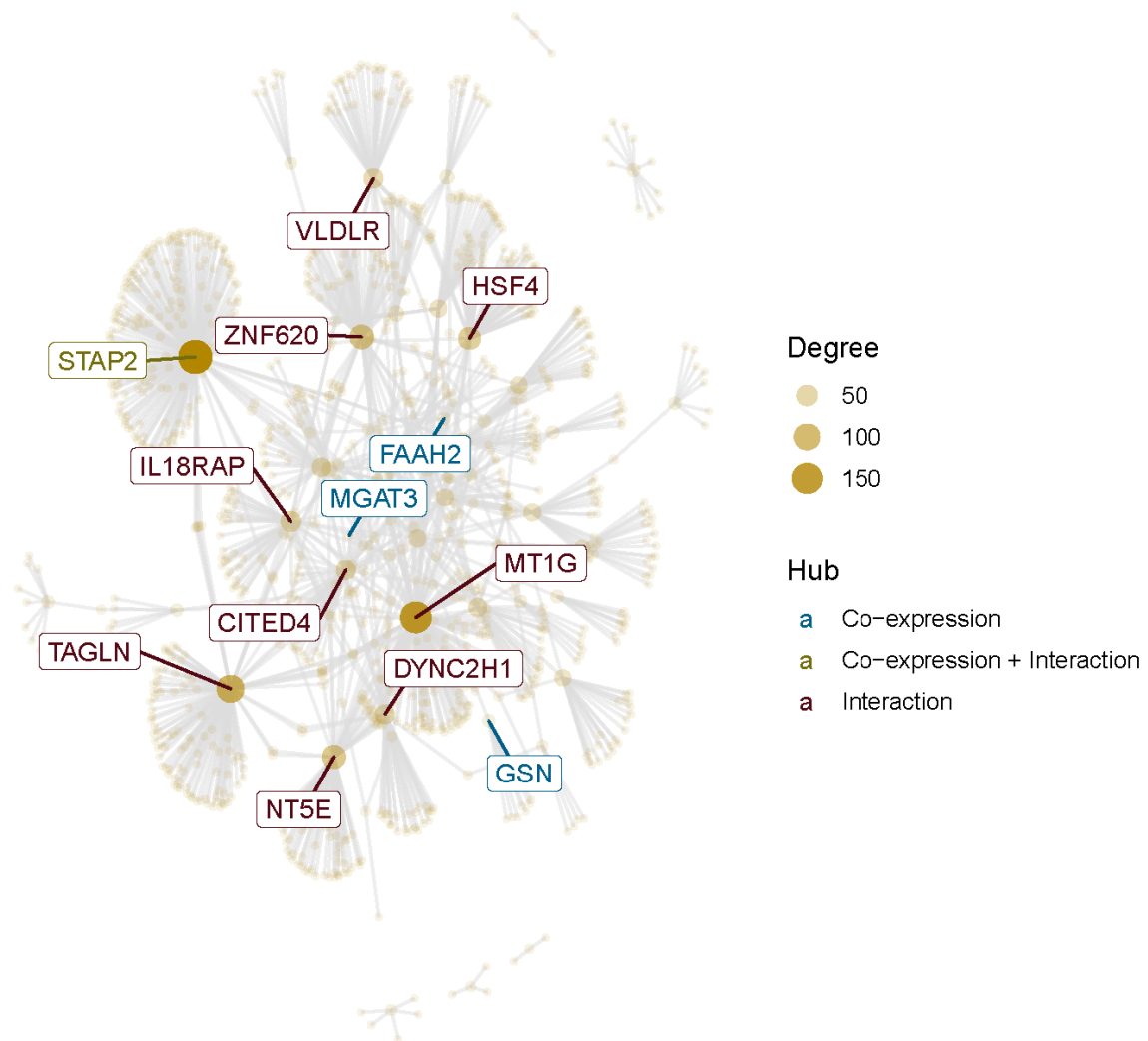
M1



**Figure 5.27: Interaction plot for M1, with gene nodes highlighted.**

The nodes represent the 142 genes of M1 plus the genes added by protein-protein interaction information. The genes are connected by co-expression and/or protein-protein interaction. Gene network of module M1 for the most connected genes (hubs) are labeled and colored based on their "origin": if originally present in the CEMiTool module, they are colored blue; if inserted from the interactions, they are colored red. The size of the node is proportional to its degree.

M3



**Figure 5.28: Interaction plot for M3, with gene nodes highlighted.**

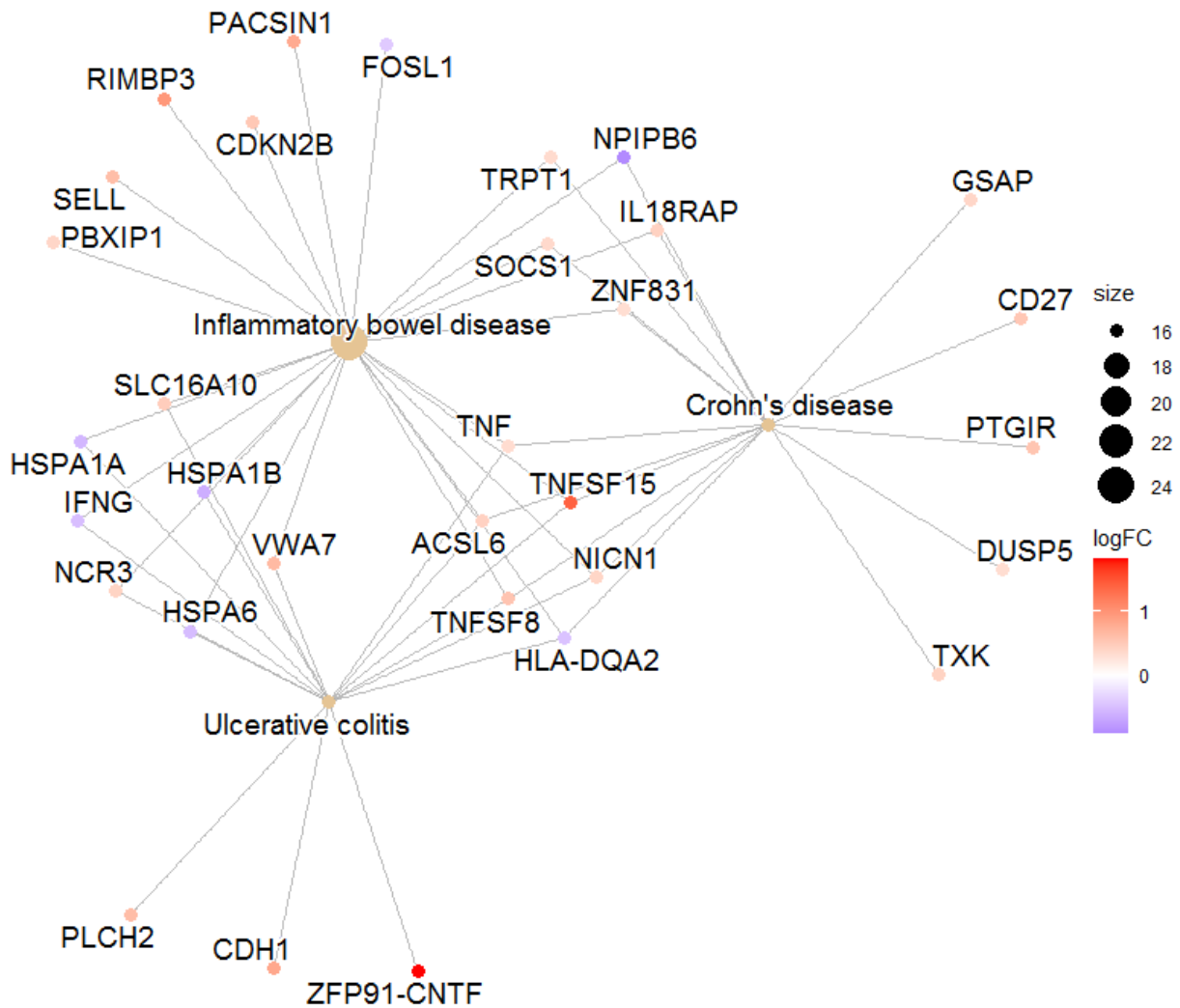
The nodes represent the 88 genes of M3 plus the genes added by protein-protein interaction information. The genes are connected by co-expression and/or protein-protein interaction. Gene network of module M1 for the most connected genes (hubs) are labeled and colored based on their "origin": if originally present in the CEMiTool module, they are colored blue; if inserted from the interactions, they are colored red; if they are present in the module and also in the inserted interactions, they are coloured green. The size of the node is proportional to its degree.

#### 5.4.6 Enrichr: GWAS annotation with DE list of FDR < 0.1

It has long been known that genetic variation between individuals can cause differences in phenotypes. These causal variants, and those which are tightly linked to the same region of the chromosome, are therefore present at higher frequency in disease than in health. Since we have associated dysregulation of ZEB2 with autoimmune diseases such as IBD, I wanted to investigate if the genetic region of these DE genes (FDR<0.1) has an association with autoimmune disease using the Genome-Wide Association Study (GWAS) 2019 database. The GWAS enrichment analysis was performed using an integrative web-based tool Enrichr that has a GWAS catalogue 2019 database to rank enriched terms. There were a total of 25 statistically enriched GWAS terms using a p-value < 0.05. Interestingly, the autoimmune diseases IBD, Ulcerative colitis (UC) and Crohn's disease were highly significantly enriched, with more than 16 genes overlapped to those GWAS terms (Figure 5.7). These candidate genes, affected when ZEB2 is lost, may contain single nucleotide polymorphisms (SNPs) within the gene body or nearby non-coding region that have been significantly associated with these autoimmune disease traits Figure 5.29.

**Table 5.13: Significantly enriched GWAS terms using a p-value < 0.05 and using ZEB2 KO Th1 EM pools DE genes with FDR < 0.1**

GWAS Term	Overlap	P-value
Inflammatory bowel disease	24/709	2.41E-04
Ulcerative colitis	16/456	1.81E-03
Psychosis proneness (hypomanic personality scale and revised social anhedonia scale)	2/5	2.25E-03
IgG glycosylation	3/23	4.96E-03
Metabolic traits	4/46	5.23E-03
Amino acid levels	2/9	7.78E-03
Pulmonary function (smoking interaction)	4/54	9.22E-03
Idiopathic osteonecrosis of the femoral head	2/10	9.62E-03
Migraine - clinic-based	2/10	9.62E-03
Blood and toenail selenium levels	2/12	1.38E-02
Lymphoma	2/13	1.62E-02
Alzheimer's disease (late onset)	4/64	1.65E-02
Common carotid intima-media thickness	2/14	1.87E-02
Chin dimples	4/74	2.65E-02
Crohn's disease	16/613	2.65E-02
Migraine without aura	2/18	3.02E-02
Cadmium levels	3/46	3.29E-02
Basal cell carcinoma	3/47	3.48E-02
Eating disorders	2/20	3.68E-02
Endometriosis	3/49	3.87E-02
Non-melanoma skin cancer	2/23	4.75E-02
Attention function in attention deficit hyperactive disorder	2/23	4.75E-02
Bone ultrasound measurement (broadband ultrasound attenuation)	2/23	4.75E-02
Lymphocyte percentage of white cells	5/129	4.78E-02
Attention deficit hyperactivity disorder	4/90	4.89E-02



**Figure 5.29: Gene-concept network plot of IBD related autoimmune diseases from Enrichr analysis using Th1 EM pools (Gene list FDR < 0.1)**

CNETplot showing Inflammatory Bowel Disease (IBD) related autoimmune diseases (orange hub modules) from Enrichr analysis using GWAS 2019 catalogue database. A  $p$ -value < 0.05 was used as the criteria for significance filtered with minimum gene set size of 5 and using the more inclusive list of DE genes from ZEB2 deleted Th1 EM pools with FDR < 0.1.

## 5.5 Discussion

This chapter investigates the comprehensive annotation of the role of ZEB2 through various bioinformatics analytical approaches. This has allowed insight into the potential function of ZEB2 in the maintenance of lineage fidelity, function, localisation and survival of Th1 cells, and suggests additional mechanisms by which loss of ZEB2 may contribute to the progression of autoimmune diseases.

It was hypothesized that the dysregulation of ZEB2 in Th1 cells may be linked to pathogenesis of Inflammatory Bowel Disease (IBD). IBD is characterized by the dysregulated immune responses to microbiota in intestinal mucosa, especially by CD4<sup>+</sup> T cell-mediated immune responses. Th1 cells are important for protecting against infectious pathogens but are also known to be one of the factors responsible for disease pathogenesis. These cells primarily produce IFN $\gamma$  and Tumor Necrosis Factor (TNF) that, respectively, activate macrophages and direct cytotoxic CD8<sup>+</sup> T cell responses, that in turn promote elimination of intracellular pathogens such as viruses and bacteria (Manetti et al., 1993). In IBD, however, Th1 cells accumulate in the intestinal tract of individuals with Crohn's disease (CD) and are directly associated with disease. This can be supported by the KEGG analysis showing enrichment of “Intestinal immune network for immunoglobulin A” genes ITGB7 and PTPRF, important for Th1 cell interactions with the intestinal epithelial barrier (Avula et al., 2012). These are also genes that were upregulated when ZEB2 is lost in a Th1 EM. It is important to note that ITGB7 is an adhesion molecule that mediates lymphocyte migration and homing to gut-associated lymphoid tissue, it is also highly associated with CD as shown in studies by Olsen et al. (2013). ITGB7 is known to interact with the cell surface adhesion molecule MADCAM1, which is normally expressed by the vascular endothelium of the gastrointestinal tract. Studies by Abramson et al. (2001) also showed ITGB7 expressing CD4<sup>+</sup> T cells were more likely to produce the Th1 cytokine IFN $\gamma$ . The association between ITGB7 and IFN $\gamma$  is contradictory in

the presence of ZEB2, since I found that ZEB2 ablated cells have upregulated ITGB7 but reduced IFN $\gamma$ . This may suggest a protective role of ZEB2 in preventing healthy Th1 EM cells from infiltrating to the gastrointestinal tract and releasing pro-inflammatory cytokine IFN $\gamma$  that could prolong inflammation, resulting in damage to the gastrointestinal tract of IBD patients.

One of the enriched KEGG pathways was “fluid shear stress and atherosclerosis” which is a chronic inflammatory disease arising from build-up of fats, cholesterol and other substances in and on the artery walls. Interestingly there have been multiple studies demonstrating the association of IBD with such heart disease (Zanoli et al., 2015, Nguyen et al., 2014, Grainge et al., 2010). In atherosclerosis, HMOX1 (encodes HO-1) may play a protective role against the progression of atherosclerosis, mainly owing to the degradation of pro-oxidant heme, the generation of anti-oxidants, biliverdin and bilirubin, and the production of vasodilator carbon monoxide (Fredenburgh et al., 2015, Ryter et al., 2006, Abraham and Kappas, 2008). Data mined from ZEB2 ChIP-seq (Section 5.4.3) suggest that HMOX1 is directly induced by ZEB2 in Th1 cells and this further confirms a protective role of ZEB2 in IBD patients, whereby cells are protected from oxidative stress and elevated levels of inflammatory cytokines such as tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) and IFN $\gamma$ , which lead to phenotypic changes in smooth muscle cells and sets into motion a series of events culminating in atherosclerosis. It is also well established that inflammation in tissues triggered by pathogens generates metabolic stress (Sun et al., 2019). Th1 cells that react to bacterial and viral infections can respond to various stressors. Induction of HMOX1 and several heat shock proteins by ZEB2 in Th1 EM cells initiate their protective function, fortifies cells against stress conditions, enhancing cell survival and reducing susceptibility to cell death, thus endowing the cells with a survival advantage in the highly variable environment of the intestine. . Several studies have demonstrated that the up-regulation of HMOX1 is associated with anti-inflammation, anti-apoptosis, and anti-proliferative properties (Xia et al., 2008). Heat shock proteins important in survival against stress responses

were also downregulated in the ZEB2 KO, whereas the pro-apoptotic gene GIMAP4 was upregulated, suggesting an inability of Th1 EM to survive with loss of ZEB2. This protective action together with that of the other heat shock proteins, are important for Th1 EM survival and function.

Chemokines are chemotactic cytokines that can control cell migratory patterns and positioning and are critical for cell development and homeostasis. The enrichment analysis carried out commonly identified disruption in cytokine secretion and chemokine receptor activities with ZEB2 ablation. As well as the disruption of IFNG, chemokines and cytokines such as CCL18, CCL3 and BMP4 were downregulated, and CXCL10, TNF, TNFSF8 and TNFSF15 were upregulated (see Figure 5.11). One of the interesting chemokines was CCL3, which has been previously shown to be released upon the induction of Th1 responses (Annunziato et al., 2000, Schaller et al., 2017). As a critical mediator of inflammation, CCL3 attracts and stimulates both antigen-presenting cells and cytotoxic cells, an accumulation of these cells was observed in rectal biopsies of patients with active inflammatory bowel disease as shown in clinical studies by Panés and Granger (1998) and Gálvez et al. (2000). If the Th1 cells express CCL3, which is indirectly induced by ZEB2, this has the potential to be a potent chemoattractant, whereby increased CCL3 facilitates the progression of tissue damage and inflammation of the gastrointestinal tract. Measurement of chemokines/cytokines in cell culture media or staining of chemokine receptor and chemokines/cytokines would validate that the loss of ZEB2 lead to disruption of cytokine secretion and chemokine receptor activities. In addition to study the change in migratory patterns, these ZEB2 deleted cells can be injected to the periphery or tail vein of an immunocompromised mice to determine if these cells home differently compare to its wildtype counterpart.

Two other interesting genes that are part of the tumour necrosis factor superfamily, TNFSF8 (encodes CD153 a ligand for CD30) and TNFSF15 (encodes TL1A) were upregulated when



ZEB2 was deleted in Th1 EM cells. These two pro-inflammatory cytokines activate CD4+T cells, antigen-presenting cells, and neutrophils. This process is thought to be involved in the pathogenesis of IBD, for example, where TNFSF15 can directly induce pro-inflammatory cytokines, including TNF $\alpha$  in T cells to exacerbate gut inflammation (Sun et al., 2008, Jin et al., 2013). If these genes are usually repressed by ZEB2, it could again suggest a protective role of ZEB2 in Th1 EM cells by reducing inflammation through repression of these cytokines. Furthermore, subsequent additional GWAS enrichment analysis also supported the association of SNPs in these genes to IBD patient populations. Concurring with the genetic findings, increases in both TNFSF8 and TNFSF15 expression have been reported in inflamed Crohn's disease lesions in patients (Prehn et al., 2004, Bamias et al., 2010). I now provide molecular insights into TNFSF8 and TNFSF15 in promoting chronic inflammation in the gut, by showing that they are normally repressed by ZEB2. Additionally, studies by Xu et al. (2014) and Deng et al. (2017) show that TNFSF15 inhibits endothelial growth. Thus, when ZEB2 is lost, TNFSF15 gene will be de-repressed and therefore can contribute to impairment of tissue repair in the gut suggesting a potential link of ZEB2 to IBD.

In the absence of ZEB2, Th1 EM cells produced diminished IFN $\gamma$  but substantially increased cytotoxic molecules. This effect of ZEB2 deficiency on IFN $\gamma$  production may be owing to increased PLAC8 gene transcription, since I have identified potential regulatory binding sites for ZEB2 in intronic regions of the PLAC8 gene. Additionally, studies by Slade et al. (2020) in mouse have shown that Plac8 is able to suppress IFN $\gamma$  mRNA and protein production after IL-12 stimulation. However, in their colitis transfer model, even though Plac8 may function to constrain IFN $\gamma$  production, their functional analyses suggests that Plac8 does not contribute to T cell-mediated inflammation *in vivo*, at least in the specific Th1-dependent disease models tested.

Gene set enrichment analysis was applied in my analysis, to evaluate whether genes were assigned to specific hallmark gene sets. Interestingly, gene sets affecting most phases of the cell cycles were negatively enriched, suggesting that ZEB2 ablated Th1 EM cells have lower cell proliferation. However, this requires further validation using cell proliferation assays. Previous studies of ZEB2 in cancer, have elucidated a similar role in which ZEB2 is necessary for mediation of cell proliferation, migration, invasion and apoptosis (Qi et al., 2012). Such a role for ZEB2 is plausible for Th1 EM cells, as they need to be able to quickly undergo clonal expansion to clear pathogens during inflammation.

Th1 EM cells with loss of ZEB2 have reduced heat shock protein expression, suggesting vulnerability to apoptosis under environmental stress. Such findings are consistent with T-cell acute lymphoblastic leukaemia (T-ALL) and acute myeloid leukemia (AML) mouse models, as demonstrated by Goossens et al. (2015) and Li et al. (2016a) respectively, where ZEB2 overexpression correlated with increased survival mediated through the IL-7-JAK-STAT signalling pathway. This complements the paradigm in cancer cells, with high expression of ZEB2, whereby cells display better self-renewal capability and better survival under environmental stress.

In this chapter I show that calcium ion signalling, and chemokine responses were interconnected and closely related and moreover, that they were both enriched with ZEB2 ablation. This is consistent with the studies of Feske et al. (2005), who observed that the primary pathway for calcium ions to enter the cell is via the CRAC (Calcium Release Activated Channel), whereby depleted cellular calcium is replenished by activation of the CRAC channel. Specifically, calcium ion-dependent dephosphorylation of the nuclear factor of activated T cells (NFAT) by calcineurin, initiates its translocation to the nucleus for regulation of various chemokine genes in T cells. Additionally, during T cell activation, intracellular relocation of calcium takes place transferring calcium ions into mitochondria, which effectively couples TCR ligation to

enhanced bioenergetics and ATP production required for clonal expansion and secretion of cytokines. Hence, altered calcium ion regulation and homeostasis in T cells by ablation of ZEB2 may severely compromise cell activation, proliferation, trafficking and effector functions, as demonstrated by the existence of various autoimmune, inflammatory and immunodeficiency syndromes (Trebak and Kinet, 2019, Partiseti et al., 1994, Le Deist et al., 1995, Feske et al., 2005). More recently, studies by Gladka et al. (2020) also confirms that ZEB2 plays a central role in regulating a transcriptional network of calcium-handling genes in the context of the heart.

Overall, the studies in this chapter have reinforced the molecular mechanisms by which ZEB2 coordinates and controls various pathways of cell adhesion, cell survival & proliferation, T cell activation, chemokine secretion & response, calcium ion signalling, migration and more. It is not just the genetic risk of some of those genes or the secretion of cytokines that might be dysregulated in IBD, it might be that ZEB2 has a crucial role in IBD pathogenesis as it controls all of these networks and pathways. If ZEB2 in Th1 EM cells is dysregulated, chronic inflammation and failure to undergo tissue repair of the endothelial cells of gastrointestinal tract would be expected to result.

## **CHAPTER 6: GENERAL DISCUSSION AND FUTURE DIRECTIONS**

## 6.1 Discussion

The CD4<sup>+</sup> T cell helper lineages have unique functional specialisation in preventing the growth and spread of various pathogens, while at the same time reducing collateral tissue damage and autoimmunity (Mahnke et al., 2013). In the Treg lineages, the regulatory phenotype is superimposed on this by FOXP3, ensuring that Treg do not acquire effector function in the steady state, and loss of control of this molecular network contributes to disease. In order to fine map the disease molecular mechanisms in CD4<sup>+</sup> T cells, the Barry lab has undertaken a FOXP3 centric transcriptomic approach to map the connections and functions of all of the genes in human Treg. Using our previously published FOXP3 ChIP studies to identify subordinate transcription factors which may be linked to subset specific functions, our lab previously identified ZEB2 as a FOXP3 target (Brown et al., 2018, Sadlon et al., 2010). Closer analysis of the ZEB2 gene body identified a region 68kb downstream of the ZEB2 transcriptional start site (TSS) in Intron 2 of the ZEB2 gene that was significantly bound by FOXP3, indicating that ZEB2 was potentially directly regulated by FOXP3. Furthermore, our lab exon array data indicated that ZEB2 was differentially expressed in Treg compared with Tconv (Sadlon et al., 2010). We subsequently confirmed direct transcriptional control of ZEB2 by FOXP3 (Brown et al., 2018), and others have shown that ZEB2 is directly induced by T-bet (Dominguez et al., 2015, Omilusik et al., 2015, van Helden et al., 2015), but as there was very little published data on the role of ZEB2 in the CD4 compartment in human, a comprehensive characterisation of ZEB2 expression in human T helper subsets and Treg subsets was undertaken. This revealed that ZEB2 is predominantly expressed in Th1 cells that have matured into an Effector Memory phenotype (Th1 EM). This fact, together with my finding that ZEB2 is repressed in Th1-like Treg cells (Treg1), indicates that ZEB2 has greater association with the effector T helper arm than the suppressor Treg arm of the human CD4<sup>+</sup> T cell compartment.

When the research presented in this PhD project was initiated, the expression of ZEB2 in the human CD4<sup>+</sup> T cell compartment remained unexplored, other than knowing it is expressed higher in a T helper compared with a Treg. Since starting this project, interest in ZEB2 in the immune system has intensified and its role in several immune cell types has been partially elucidated, although until now there has been little published about the role of ZEB2 in CD4<sup>+</sup> T cells. I have used healthy donor samples to interrogate the lineage and maturation restriction of ZEB2 in order to define its role, and this revealed that ZEB2 is highly expressed in Th1 cells. This raised the possibility that ZEB2 expression is controlled by the Th1 lineage defining transcription factor T-bet, as has been proposed in other cells in mouse. Omilusik et al. (2015), Dominguez et al. (2015) and van Helden et al. (2015) showed that T-bet directly drives expression of ZEB2 in mouse CD8<sup>+</sup> T cells and NK cells, and our expression profile showing that T-bet and ZEB2 are expressed predominantly in the same CD4 Th1 population, supported the likely co regulation of ZEB2 by T-bet in human.

The published studies examining ZEB2 expression in human Th1 are contradictory. Studies from Schmiedel et al. (2018) showed similar enrichment of ZEB2 in the Th1 cell populations in agreement with my findings. However, studies from Höllbacher et al. (2020) showed as there was no enrichment of ZEB2 expression in their T helper populations. This disparity may be owing to the CD25 MACS enrichment strategy that was carried out in the Höllbacher study which excludes CD25<sup>+</sup> Th cells from the Th pools generated, and, similar to my initial T helper population purification strategy (see Section 3.4.4), where I discovered that Th1 EM would be underrepresented by this approach, and even potentially included in the Treg pool based on CD25 enrichment, the results are inconclusive. The CD25 expression on Th1 EM might suggest that Th1 EM cells require ZEB2 expression *in vivo* or in the periphery, as these cells are in a pre-activated state (CD25<sup>hi</sup>) ready to perform an effector function when encountering intracellular pathogens. The significance of these findings remains to be

discerned, but a closer examination of subset or populations with a better ZEB2 monoclonal antibody that is sensitive enough to measure the low abundance of ZEB2 protein expression overall in human CD4<sup>+</sup> T cells will benefit the field.

Having demonstrated that ZEB2 was co expressed with T-bet in Th1 cells, the deep delve into all Th subsets in the human CD4 pool revealed the observation that ZEB2 is not expressed robustly in the strongly T-bet positive Th1/17 cells, suggesting that ZEB2 is not controlled solely by T-bet, as there are high levels of T-bet in 2 other lineages (Treg1, Treg 1/17) but little or no ZEB2 expression in these cells. The new knowledge is that it is not a simple binary, T-bet controlled transcriptional regulation of ZEB2, and that it is possible that other mechanisms, possibly another transcription factors or microRNAs, inhibit ZEB2 expression in Th1/17cells. Whereas in the case of Treg1 and Treg17, I inferred that ZEB2 is strongly suppressed by FOXP3 and miR-155 using a feedforward mechanism operating in all Treg lineages in the steady state and thus FOXP3 predominates over T-bet to prevent ZEB2 activation in all Treg.

In exploring whether the role of ZEB2 is to shape maturation, lineage fidelity or stability, we explored the restriction of ZEB2 expression as CD4 subsets mature down memory phenotypes. Interestingly, when we segregated the population into CM and EM cells with CD62L, ZEB2 was almost exclusively expressed in the Th1 EM. It is proposed that Th1 EM are more differentiated cells than Th1 CM, and this may represent lineage stability, as plasticity is progressively reduced upon T cell differentiation (Geginat et al., 2014, Sallusto et al., 2018). A correlation of ZEB2 expression with degree of differentiation has also been reported in the CD8<sup>+</sup> T cell compartment where T-bet is required for cytotoxic effector function, but it should be noted that the CD8<sup>+</sup> T cell compartment does not segregate into the equivalent helper subsets seen in CD4 T cells, so may not require subordinate transcription factors for function in the same way as the CD4<sup>+</sup> T cell compartment.

Interestingly, the presence of ZEB2 was more recently reported in a cytotoxic CD4<sup>+</sup> (CD4-CTL) cluster, described predominantly in studies where scRNA-seq analysis was performed on the CD4-TEMRA population and SARS-CoV-2 reactive CD4<sup>+</sup> T cells (Patil et al., 2018, Meckiff et al., 2020). Studies by Serroukh et al. (2018) sorted out the CD4-CTL and carried out RNA-seq analysis to compare with Th1 CM and they observed high ZEB2 expression exclusively found in CD4-CTL, but when these cells were activated *in vitro*, ZEB2 expression was no longer specifically expressed in the CD4-CTL. This study was consistent with my findings that when T helper cells were activated, ZEB2 expression was altered and was no longer uniquely expressed in Th1 EM cells. I also observed that ZEB2 deletion in Th1 EM cells resulted in the expression of various cytotoxic molecules such as GZMA, GZMH and GZMK, consistent with its role in repressing these genes in steady state Th1 cells. In addition, a very recent publication from Krueger et al. (2021) also suggested that Salmonella-induced CX3CR1<sup>+</sup> Th1 cells are cytotoxic and depend on the transcription factor ZEB2. However, my analysis of RNA-seq between Th1 CM and Th1 EM showed no enrichment of key CD4-CTL signature genes such as CRTAM, EOMES, PRF1, GNLY and GZMB. Hence, suggesting that in the steady state ZEB2 is required to maintain the function of Th1 EM cells and restrict cytotoxic function.

Whether ZEB2 has a different role in Treg requires further investigation. Detection of ZEB2 expression in Treg Effector Memory cells was shown to be at a similar level compared with Tconv Effector Memory cells. On the other hand, FOXP3 in Treg is known to positively regulate miR-155 and then the coordinated action of FOXP3 and miR-155 blocks possible effector function of ZEB2 (Brown et al., 2018, Sadlon et al., 2010). Therefore, the expression of ZEB2 in a Treg remains controversial. The presence of ZEB2 in Treg Effector Memory cells raises the possibility that ZEB2 may have a different role in a Treg compared with a Th1 EM. Consistent with this, enforced expression of ZEB2 in Treg and Tconv resulted in different



effects on IL-10 expression. ZEB2 enforced expression in Treg resulted in increased IL-10 expression, whereas ZEB2 overexpression in Tconv resulted in reduced expression of this anti-inflammatory cytokine. Perhaps transient ZEB2 expression may be induced to prevent the Treg Effector Memory from carrying out its suppressive or regulatory functions if recruited to a site of active pathogen challenge, and even support migration and homing, while they are recirculating between peripheral blood and non-lymphoid tissues. However, if suppressive activity is required in the tissue, FOXP3 may still tightly repress ZEB2 so that anti-inflammatory cytokine, IL-10 can be secreted to modulate an immune response. This may also be a mechanism by which a Treg can terminate an effector response once the pathogen is cleared.

The results presented here suggest that while ZEB2 is crucial for the effector function of Th1 EM cells, aberrant expression of ZEB2 in both Tconv as well as Treg may impair their normal function. Therefore, tight regulation of ZEB2 is required, not only in Treg where ZEB2 is normally kept repressed, but in Tconv as well, where ZEB2 is normally highly restricted to the Th1 EM cells. These findings have opened interesting avenues of research, raising many questions regarding what mechanisms are involved for ZEB2 to escape FOXP3 and miR-155 tight repression and if ZEB2 regulates a different set of genes in a Treg compared to a Tconv. Thus, the functional relevance of ZEB2 in a Treg requires further examination.

Here, my data supports my finding that ZEB2 is almost exclusively expressed in Th1 EM cells and is downstream of T-bet. This places ZEB2 as key in maintaining Th1 function and fidelity as the classical Th1 cytokine, IFN $\gamma$ , is lost with knockout of ZEB2. Furthermore, ZEB2 in Th1 EM is implicated in restricting gain of function that is predicted to be pathogenic. Evidence for this is provided by an increase in cytotoxic molecules, signature genes that are associated with T follicular helpers, Th2 markers with increased CCR4, Th17 markers with increased CXCR6 and other genes that are not Th1 markers, but are characteristic of a number of other cell

lineages or cell types. In addition, consistent with a role for ZEB2 in controlling cell morphology and physical properties in EMT, it also regulates cell adhesion properties in Th1EM cells which was observed to be significant by enriched KEGG pathway terms (Figure 5.8 & Table 5.4). All these suggest that ZEB2 is important in ensuring correct Th1 cell migration and trafficking to the right location to avoid non-specific or inappropriate effector response or effector responses in the wrong place. Therefore, ZEB2 is essential for the maintenance of Th1 EM cell identities and immune function.

The generation of Th1 cells from naïve cells is crucial for host protection. The implications of ZEB2 expression being restricted to Th1 EM are significant, as Th1 are a major effector T cell subtype, mainly producing Th1 cytokines such as IFN $\gamma$  and are characterized by the key transcription factor T-bet. Following activation with viral or intracellular bacterial antigen, naïve CD4<sup>+</sup> T cells undergo proliferation and differentiation into a Th1 helper lineage population to allow for a type 1 immune response. I show that ZEB2 doesn't seem to play a role in the differentiation of Th1 cells, but it greatly impacts the ability of Th1 cells to produce Th1 pro-inflammatory cytokine IFN $\gamma$ . It is believed that an overexuberant or inappropriate Th1 response is involved in the pathogenesis of multiple autoimmune diseases including Type 1 diabetes mellitus, Rheumatoid arthritis, SLE and Inflammatory bowel disease. A number of studies have shown that ZEB2 promotes effector function, and if that is dysregulated as part of these diseases, it could therefore be a therapeutic target to mitigate these diseases. In order to better understand the transcriptional program controlled by ZEB2 in Th1 EM cells, I have used CRISPR/Cas9 to delete it in primary human CD4<sup>+</sup> T cells, and then I explored the consequences. When I deleted ZEB2 from Th1 EM, its loss led to the reduction of IFN $\gamma$  expression. As this is a signature Th1 effector gene, this suggested a role for ZEB2 in Th1 effector function and lineage fidelity. Mechanistically, this could be by direct or indirect regulation of the IFN $\gamma$  gene, as it is possible that the reduction of IFN $\gamma$  was owing to the

upregulation of PLAC8 which is directly regulated by ZEB2, based on my findings showing it to be a ZEB2 ChIP target in an activated Th1 cell. Consistent with this, studies by Slade et al. (2020) also reported that PLAC8 was important in suppressing IFN $\gamma$  mRNA and protein production.

In addition, the work performed for this PhD project found that ZEB2 also regulates a number of Heat Shock Protein (HSP) that have not been associated with ZEB2 function before. Importantly, these HSPs were also a target of ZEB2 ChIP, suggesting that ZEB2 induction of these HSP proteins in Th1 EM cells is important in regulating immune responses by direct or indirect mechanisms. HSPs can be secreted by Th1 cells in response to exposure to environmental stress. Recently, increasing attention has been devoted to the role of HSPs in immune processes and several lines of evidence link the expression of HSPs to the development and prognosis of IBD. My analysis of ZEB2 ChIP-seq data shows that HSP70 (encoded by HSPA1A, HSPA5 & HSPA1B) can be directly induced by ZEB2 and elevated expression of HSP70 has been reported in patients with intestinal inflammatory diseases compared with healthy individuals (Samborski and Grzymisławski, 2015). Additionally, several studies by Tanaka et al. (2007) and Borges et al. (2012) have also pointed out that HSP70 can trigger IL-10 production resulting in anti-inflammatory effects in intestinal inflammation in a mouse model, suggesting a protective effect. It is possible that in a highly inflamed site of an IBD patient, ZEB2 tries to upregulate HSP70 proteins to trigger IL-10 to inhibit further inflammation. Taken together, these observations denote that ZEB2 can disrupt the physiological induction of HSPs and, hence, influence the intestinal vulnerability to infection and inflammation. Production of HSPs in Th1 EM cells, perhaps potentiated by ZEB2, fortifies cells against stress conditions, enhancing cell survival and reducing susceptibility to cell death. Broadly protective function of HSPs regulated by ZEB2, reflects their ability to suppress several forms of cell death, including apoptosis. Such enhanced survival capability by ZEB2 has

previously been reported in leukaemia cells in studies by Goossens et al. (2015) and Li et al. (2016a). In this work, I have shed the light on the role of ZEB2 in inducing HSPs, which in the context of IBD may prevent further disease progression.

This PhD project has also shown that dysregulation of ZEB2 function in Th1 cells can potentially explain some of the pathology in multiple autoimmune diseases, including, importantly, Inflammatory Bowel Disease (IBD) pathogenesis, which has been discussed extensively in Chapter 5. The extensive meta-analysis of the ZEB2 knockout data revealed strong links to IBD associated gene sets, KEGG pathway and GWAS, consistent with the major cause of diseases being driven by the disruption of immune homeostasis in Th1 cells of the gastrointestinal tract. One of the suggested protective roles of ZEB2 is in preventing Th1 EM cells from infiltrating the gastrointestinal tract and releasing pro-inflammatory cytokines that could prolong inflammation, resulting in damage to the gastrointestinal tract of IBD patients. Clinical trials of anti-IFN $\gamma$  (Fontolizumab) (Reinisch et al., 2010, Hommes et al., 2006) therapy for IBD treatment did not achieve the primary end point. Therefore, it is more complex than simply regulating IFN $\gamma$ , and the role of ZEB2 is in controlling a broad set of genes to shape effector function. As such, ZEB2 is not only a useful marker for disease severity of IBD, it also has potential as a therapeutic drug target.

Although this study of ZEB2 in CD4<sup>+</sup> T cell is presented without a clinical cohort, the rigorous bioinformatics pathway analysis indicated novel functions of ZEB2 that indicate significant impact on disease pathogenesis of autoimmunity, if dysregulated. Therefore, further investigation is necessary as suggested below to prove the clinical implications of ZEB2 in autoimmune disease such as IBD.

## 6.2 Future Direction

In my PhD project, I have uncovered, for the first time, a novel role for ZEB2 in CD4<sup>+</sup> T cells and its association with IBD pathogenesis. Characterisation of ZEB2 expression by mRNA quantitation also identified Th1 Effector Memory cells as the highest ZEB2 expressing cells. However, such methodologies are limited by the number of purified subsets or populations for ZEB2 expression screening, so increasing the sample size and including disease samples would enhance the clinical importance of the findings. Further studies are warranted to fully characterise ZEB2 at the protein level using a better ZEB2 monoclonal antibody that are sensitive and specific enough to measure the relatively low ZEB2 protein expression overall in human primary T cells at the single cell level. This ZEB2 antibody could then be coupled with phenotypic markers to fully capture other rare subsets that might be missed using our memory chemokine receptor panel, such as T follicular helpers and cytotoxic T cells, allowing for a comprehensive phenotypic analysis of ZEB2.

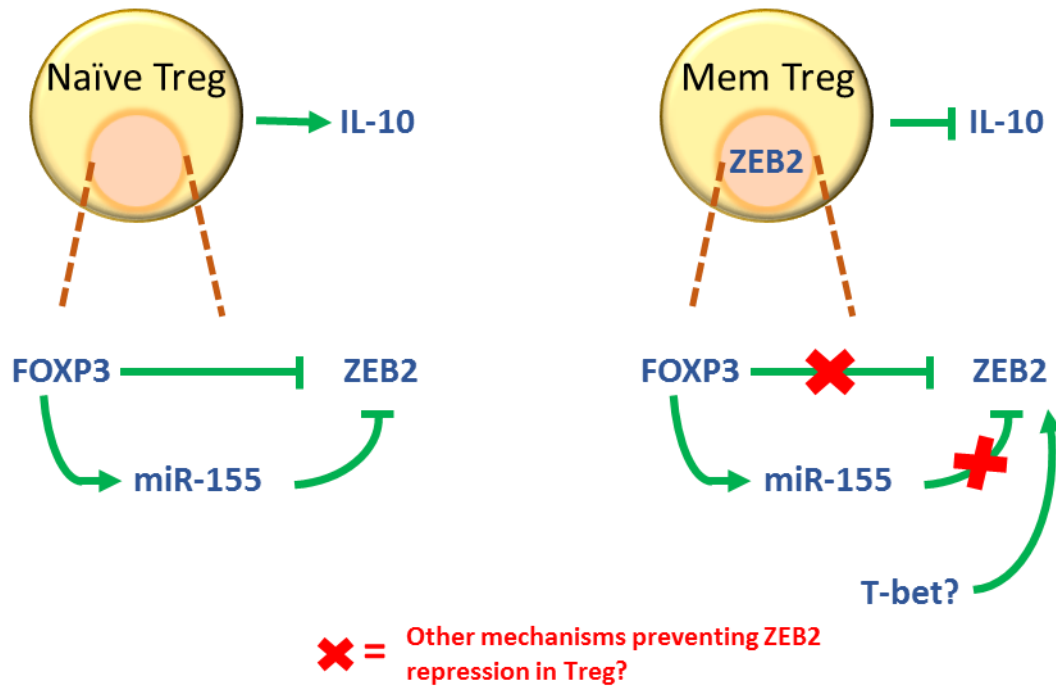
To further validate the findings of the meta-analysis, some functional studies such as cytotoxic assays can be carried out with ZEB2 knocked out Th1 EM cells in conjunction with other well-known cytotoxic cells CD8 T cells and CD4-CTL to compare its cytotoxic levels. In parallel, we could also conduct further validation on the RNA-seq data to determine if the changes in gene expression is similar at the protein level by either staining with an antibody for flow cytometry or western blot.

In addition to using a ZEB2 antibody to phenotype protein expression in CD4<sup>+</sup> T cell subsets, a versatile antibody could also be used to carry out a stringent ZEB2 ChIP-seq in human CD4<sup>+</sup> T cells. ChIP-seq is a powerful method for identifying genome-wide DNA binding sites for transcription factors like ZEB2. Currently, the closest ChIP-seq data for ZEB2 in immune cells was in mouse CD8<sup>+</sup> T cells by (Omilusik et al., 2015), therefore no ZEB2 ChIP-seq has been carried out in human CD4<sup>+</sup> T cells or even in the well characterised human CD8<sup>+</sup> T cells.

Having human CD4<sup>+</sup> T cell ZEB2 ChIP-seq data would be more accurate compared with ChIP-seq data from non-immune cell origin, such as HEK293 (eGFP tagged ZEB2) and K562 cell lines from the ENCODE project as shown in this thesis. Without a working human ZEB2 antibody, using a His-tagged ZEB2 transgene in a T cell pool and an anti His ChIP antibody could also be an alternative in determining the direct binding of ZEB2 in human CD4<sup>+</sup> T cells.

Interestingly, while the majority of my PhD studies focussed on the role of ZEB2 in the Tconv and specifically the Th1 Effector Memory cells, the detectable expression of ZEB2 in the memory Treg also gained my attention, as ZEB2 was able to escape from the FOXP3 and miR-155 feed forward repression seen in the naïve Treg pool. However, owing to the time constraints of this PhD project, these exciting preliminary findings were not examined further. To further determine if ZEB2 plays a different role in Treg and Tconv, RNA-seq could be performed on cells overexpressing ZEB2 in both Treg and Tconv. This would capture the global transcriptome profile differences between ZEB2 overexpressing and wildtype cells using deep-sequencing technologies. Differentially expressed genes from both subsets could later be intersected to further elucidate if ZEB2 has overlapping or non-redundant roles in Treg and Tconv. My proposed model of ZEB2 function and regulation in the Treg is shown in Figure 6.1. It is possible that other, as yet undiscovered, mechanisms could disrupt the repression of ZEB2 in the memory Tregs. In this scenario, ZEB2 may be directly induced by T-bet or perhaps de-repressed by loss of miR-induced repression. The presence of T-bet might also result in other mechanisms that may inadvertently disrupt FOXP3 repression. High ZEB2 expression in a memory Treg results in the loss of IL-10 expression causing dampening of Treg suppressive ability. This requires experimental confirmation by testing Treg overexpressing ZEB2 in mixed lymphocyte reaction (MLR) suppression assays. This may potentially be to reduce suppression whilst cells are circulating between blood and non-lymphoid tissues. It is possible that at the site of inflammation, after pathogen clearance, ZEB2 would then be switched off by FOXP3 to

allow for Treg suppressive function to resume. T-bet expression in a FOXP3-expressing Treg, does not result in a pro-inflammatory function but instead, the presence of T-bet tunes the Treg to respond to a Th1 inflammatory cue to limit the inflammatory response and its expression is more so to support the cells to correctly respond to the right environmental cues to the site of inflammation and carry out its function.



### Figure 6.1: ZEB2 role and regulation in Treg paradigm

Model suggesting that in a naïve Treg where FOXP3 is high, ZEB2 is tightly repressed by FOXP3 and miR-155 in a feed forward loop, as proposed by Brown et al. (2018), to allow for production of IL10 for suppression of an immune response. In activated Central Memory or Effector Memory Treg cells, Treg express ZEB2 even in the presence of FOXP3, reducing mediators of suppressor function. It is possible that there may be other mechanisms yet to be discovered that could disrupt the repression of ZEB2, perhaps T-bet might directly induce ZEB2 in a Treg.

Further clinical studies could be performed on PBMC of IBD samples to determine if ZEB2 is aberrantly expressed in Th1 EM. It is possible that in IBD samples, ZEB2 may no longer exhibit lineage restriction but instead be expressed highly in all T helper cells. In this scenario, the cells may constantly receive signals to migrate and home to the gastrointestinal tract to carry out effector function. Additionally, it is also possible that Th1/17 cells normally expressing T-bet, but not ZEB2, may initiate ZEB2 expression in the disease cohort. Therefore, ZEB2 could have

utility as a potential marker for IBD prognosis or disease severity by comprehensive screening of ZEB2 expression across the helper lineage populations.

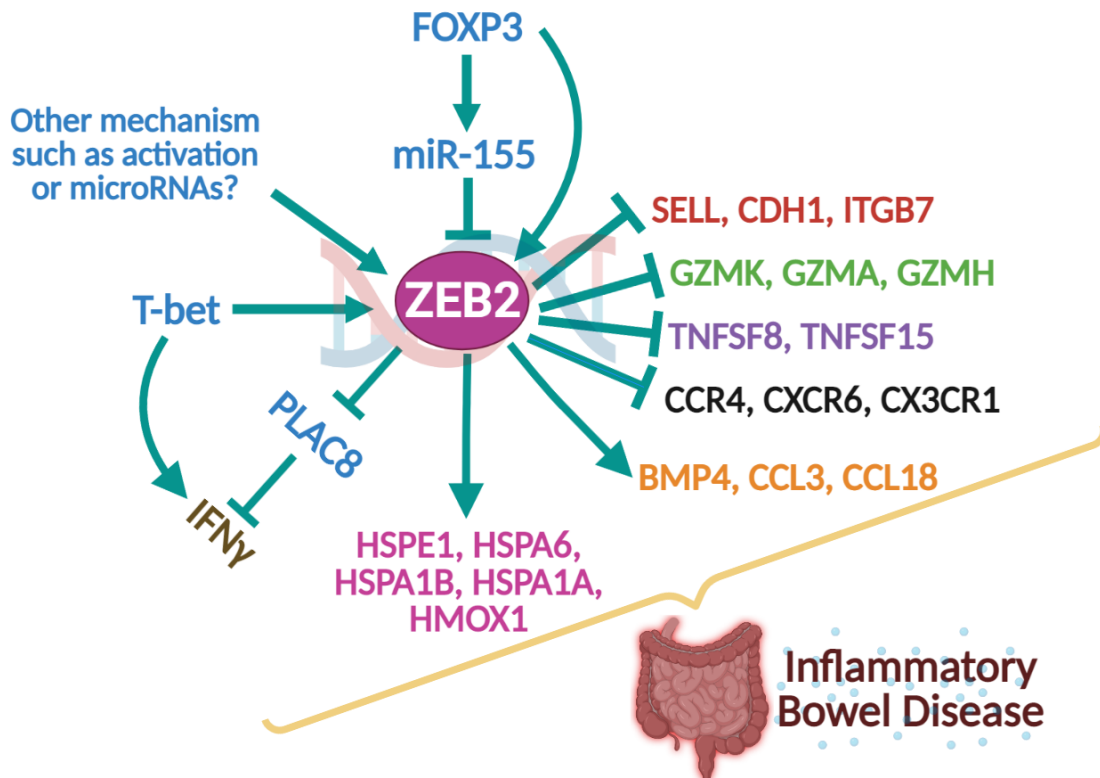
Although, this project has been focused only on human samples, mouse models may help in answering several important questions including the origin and fate of ZEB2 expressing cells. Although the mouse model does not necessarily translate well to human, ZEB2 knockout in CD4<sup>+</sup> T cells mice may reveal the significance of ZEB2 in CD4<sup>+</sup> T cells. For example, will the loss of ZEB2 in mouse CD4<sup>+</sup> T cells result in the pathogenesis of IBD? A mouse model will be beneficial in providing more evidence to support my thesis findings of the ZEB2 association to IBD.

### 6.3 Conclusions

This PhD project aimed to perform an extensive analysis of ZEB2 to examine its function and role in the CD4<sup>+</sup> T cell compartment. Based on the results from this study, it is clear that ZEB2 has a prominent role in Th1 EM cells. The findings from this PhD project provide insight into the crucial role of ZEB2 in the CD4<sup>+</sup> T cell and its importance in maintaining Th1 lineage fidelity and function. For example, ZEB2 regulates a small but significant number of EM and CM genes that are important for Th1 Effector Memory differentiation and function. Moreover, ZEB2 can control TNF signalling via the NF- $\kappa$ B pathway to prevent chronic inflammation and allow for tissue repair. ZEB2 showed anti-proliferative signalling, suggesting the importance of ZEB2 in Th1 cell proliferation and expansion in response to foreign pathogens. ZEB2 also repressed chemokine receptors CCR4, CXCR6, CXCR1 to prevent Th1 cells from responding to chemokines and the cascade of calcium ion signalling resulting in incorrect homing of cells to non-inflamed sites. Genes related to protection against cellular stress responses were also upregulated by ZEB2 to allow for better survival against apoptosis. Therefore, if ZEB2 in Th1 EM cells is dysregulated, it may severely compromise signalling pathways altering cell activation, proliferation, trafficking and effector functions, resulting in IBD (Figure 6.2). In



conclusion, ZEB2 expression in Th1 EM cells serves as a lineage fidelity checkpoint for IBD pathogenesis. Further functional and clinical studies as mentioned above may reveal the full characteristics of ZEB2 and may provide future directions in using ZEB2 as a potential biomarker in therapeutics/targeted therapy.



**Figure 6.2: ZEB2 regulation and function in CD4+ T cell in implicating IBD pathogenesis**  
 Model showing mechanisms regulating ZEB2 expression and the function of its downstream target genes including adhesion and motility (red), preventing cytotoxic function (green), maintaining organ homeostasis and initiating tissue responses (purple), localisation to the site of inflammation via chemokine receptors and calcium signalling (black & orange), enhancing cell survival under stress environment (pink), maintaining Th1 fidelity and functions (brown), each of which is known to have implications in disease causality of Inflammatory Bowel Disease.

**CHAPTER 7: APPENDIX**

## 7.1 RNA-seq Libraries Preparation Summaries

**Table 7.1: RNA-seq libraries preparation summary of ZEB2 knockout in Th1 EM clones from donors #070618 & #150618**

Library no.	Sample	Qubit RNA Input (ng)	Adaptor Dilution	NEB Index	Index Sequence		PCR cycles	DNA 1k Experion			Kappa Quant. Conc. (nM)	Pooling (20nM)		
					i5 Barcode	i7 Barcode		Conc. (ng/μL)	Molarity (nmol)	Fragment Size bp		Library Vol. (μL)	1x TE (μL)	Pooling Vol. (μL)
1	07 WT1 clone	16.9	5 fold	3	universal	GCCTAA	11	9.10	41.70	437.33	188.57	2	16.9	2.5
2	07 KO6 clone	8.22	25 fold	4	universal	TGGTCA	11	11.30	51.20	438.70	209.09	2	18.9	2.5
3	07 KO9 clone	6.3	5 fold	5	universal	CACTGT	11	13.50	58.30	461.19	126.39	2	10.6	2.5
4	07 KO12 clone	34.8	5 fold	6	universal	ATTGGC	11	8.10	33.30	459.68	208.80	2	18.9	2.5
5	15 WT11 clone	22.6	5 fold	7	universal	GATCTG	11	15.90	74.10	467.36	150.71	2	13.1	2.5
6	15 WT21 clone	27.2	5 fold	8	universal	TCAAGT	11	20.80	98.90	418.89	176.63	2	15.7	2.5
7	15 WT22 clone	23.8	5 fold	9	universal	CTGATC	11	7.70	33.60	429.50	190.84	2	17.1	2.5
8	15 KO1 clone	37.6	5 fold	10	universal	AAGCTA	11	10.80	50.10	426.61	185.68	2	16.6	2.5

**Table 7.2: RNA-seq libraries preparation summary of ZEB2 knockout in Th1 CM & EM pools from donors #08, #10, #16 & #18.**

Library no.	Sample	Qubit RNA Input (ng)	Adaptor Dilution	NEB Index	Index Sequence		PCR cycles	DNA 1k Experion			Kappa Quant. Conc. (nM)	Pooling (20nM)		
					i5 Barcode	i7 Barcode		Conc. (ng/μL)	Molarity (nmol)	Fragment Size bp		Library Vol. (μL)	1x TE (μL)	Pooling Vol. (μL)
1	08 Th1 CM WT	200.00	25-fold	1	universal	CGTGAT	11	28.4	112.9	381.14	135.00	2	11.5	2
2	08 Th1 CM KO1	186.00	25-fold	2	universal	ACATCG	11	29.7	125.9	357.43	180.68	2	16.1	2
3	08 Th1 CM KO2	200.00	25-fold	4	universal	TGGTCA	11	28.6	122.3	354.32	182.97	2	16.3	2
4	10 Th1 CM WT	79.46	100-fold	5	universal	CACTGT	12	4.5	17	401.07	50.18	2	3	2
5	10 Th1 CM KO1	200.00	25-fold	6	universal	ATTGGC	11	23	91	382.95	149.66	2	13	2
6	10 Th1 CM KO2	200.88	25-fold	7	universal	GATCTG	11	26.7	111.9	361.52	176.00	2	15.6	2
7	16 Th1 CM WT	200.00	25-fold	9	universal	CTGATC	11	36	140.2	389.05	255.80	2	23.6	2
8	16 Th1 CM KO	200.00	25-fold	10	universal	AAGCTA	11	40	169.1	358.40	197.15	2	17.7	2
9	08 Th1 EM WT	68.85	100-fold	11	universal	GTAGCC	12	6.8	28.1	366.66	35.23	2.3	1.7	2
10	08 Th1 EM KO1	59.95	100-fold	12	universal	TACAAG	13	10.8	47.5	344.50	64.93	2	4.5	2
11	08 Th1 EM KO2	90.60	100-fold	13	universal	TTGACT	12	6	25.1	362.19	38.65	2.1	1.9	2
12	10 Th1 EM WT	36.16	100-fold	14	universal	GGAACT	13	3	11.9	381.97	27.80	2.9	1.1	2
13	10 Th1 EM KO1	39.30	100-fold	15	universal	TGACAT	13	3.4	13.8	373.30	26.60	3	1	2
14	10 Th1 EM KO2	41.58	100-fold	16	universal	GGACGG	13	4.2	17.9	355.51	30.63	2.6	1.4	2
15	16 Th1 EM WT	200.00	25-fold	18	universal	GCGGAC	11	26.6	111.8	360.49	137.97	2	11.8	2
16	16 Th1 EM KO	200.00	25-fold	19	universal	TTTCAC	11	22.6	94.1	363.89	114.85	2	10.1	2.1
17	18 Th1 EM WT	4	100-fold	1	universal	CGTGAT	13	1.20	5.00	378.33	25.49	3.9	1.1	2.5
18	18 Th1 EM KO	2.3	100-fold	2	universal	ACATCG	13	1.90	7.80	401.93	34.77	2.9	2.1	2.5

## 7.2 Next Generation Sequencing (NGS) Reports

### 7.2.1 ngsReports: ZEB2 Knockout in Th1 EM Clones

Vincent Wong

- FastQC Summary
  - Read Totals
  - FastQC Summary
  - Per Base Sequence Quality
  - Per Sequence Quality Scores
  - Per Base Sequence Content
  - Per Sequence GC Content
  - Sequence Length Distribution
  - Sequence Duplication Levels
  - Adapter Content
- Session Information

## FastQC Summary

Show  entries

Search:

Summary statistics for all libraries

Filename	Total	Flagged As Poor Quality	Sequence Length	%GC	File Type	Encoding
KO1_combined_R1.fastq.gz	58,075,730	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
KO1_combined_R2.fastq.gz	58,075,730	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
KO12_combined_R1.fastq.gz	56,434,907	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
KO12_combined_R2.fastq.gz	56,434,907	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
KO6_combined_R1.fastq.gz	52,006,780	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
KO6_combined_R2.fastq.gz	52,006,780	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
KO9_combined_R1.fastq.gz	44,724,668	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
KO9_combined_R2.fastq.gz	44,724,668	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT1_combined_R1.fastq.gz	35,519,040	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9

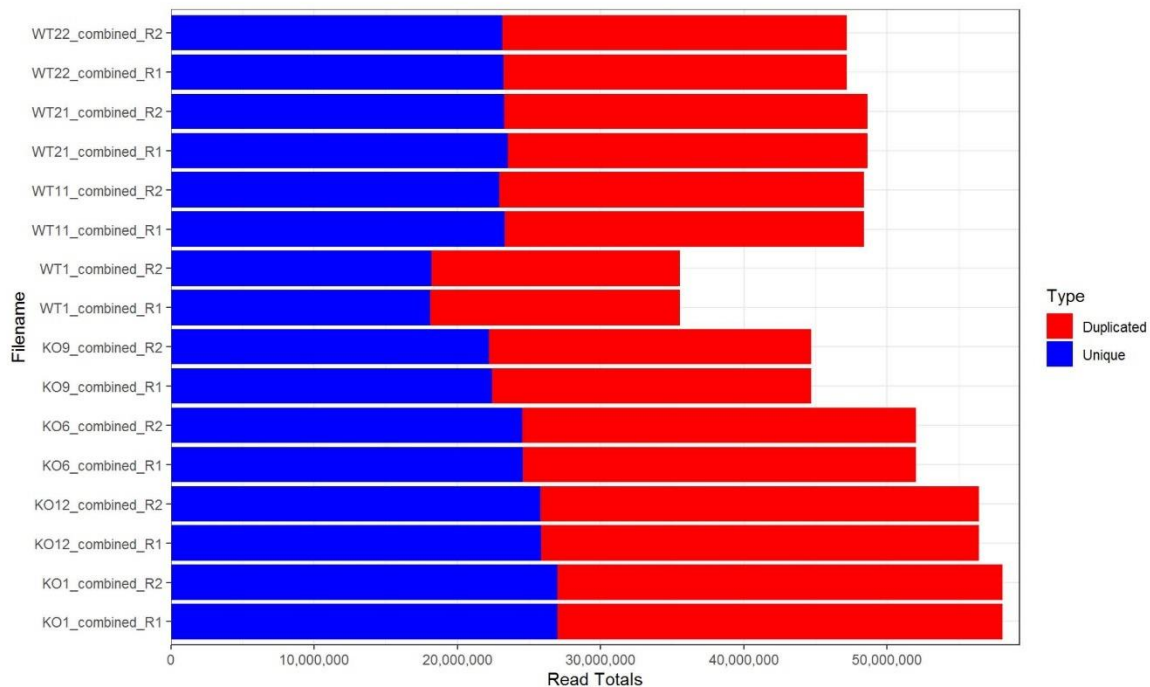
Filename	Total	Flagged As Poor Quality	Sequence Length	%GC	File Type	Encoding
WT1_combined_R2.fastq.gz	35,519,040	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT11_combined_R1.fastq.gz	48,392,013	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT11_combined_R2.fastq.gz	48,392,013	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT21_combined_R1.fastq.gz	48,648,321	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT21_combined_R2.fastq.gz	48,648,321	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT22_combined_R1.fastq.gz	47,186,653	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
WT22_combined_R2.fastq.gz	47,186,653	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9

Showing 1 to 16 of 16 entries

Previous 1 Next

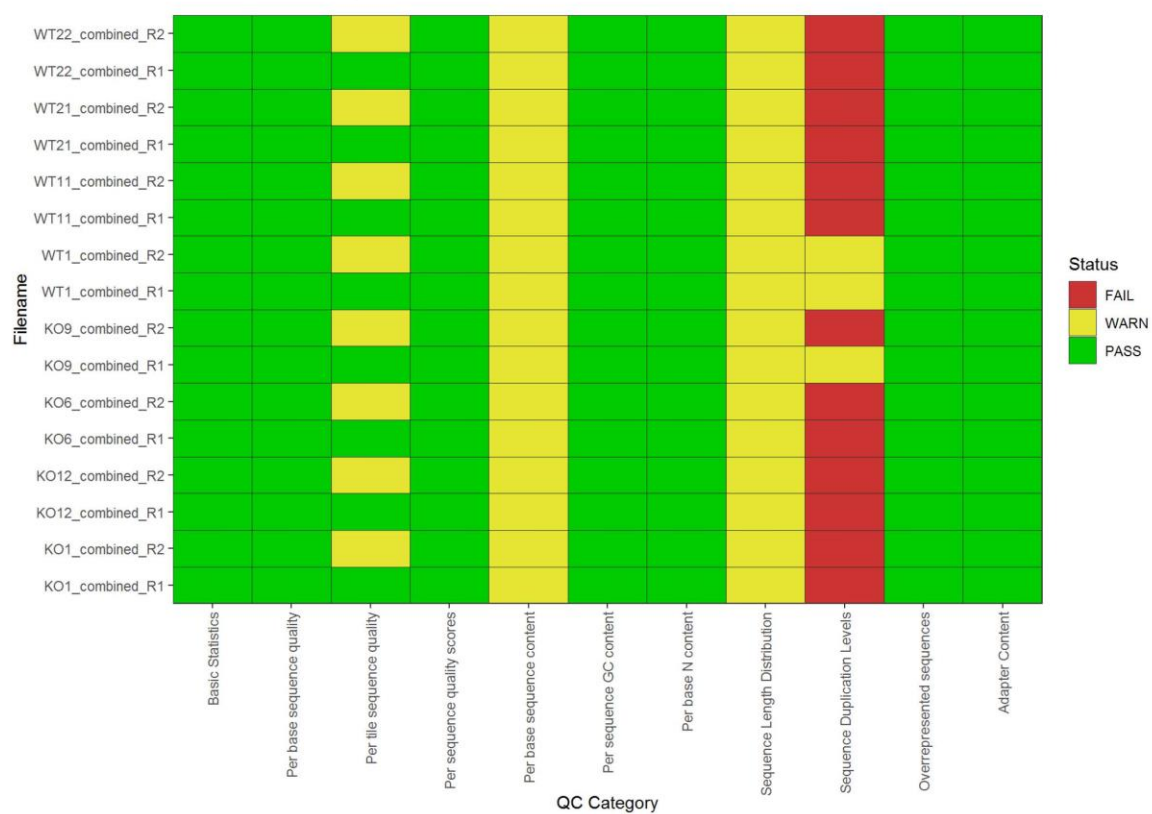
## Read Totals

Library Sizes ranged between 35,519,040 and 58,075,730 reads.



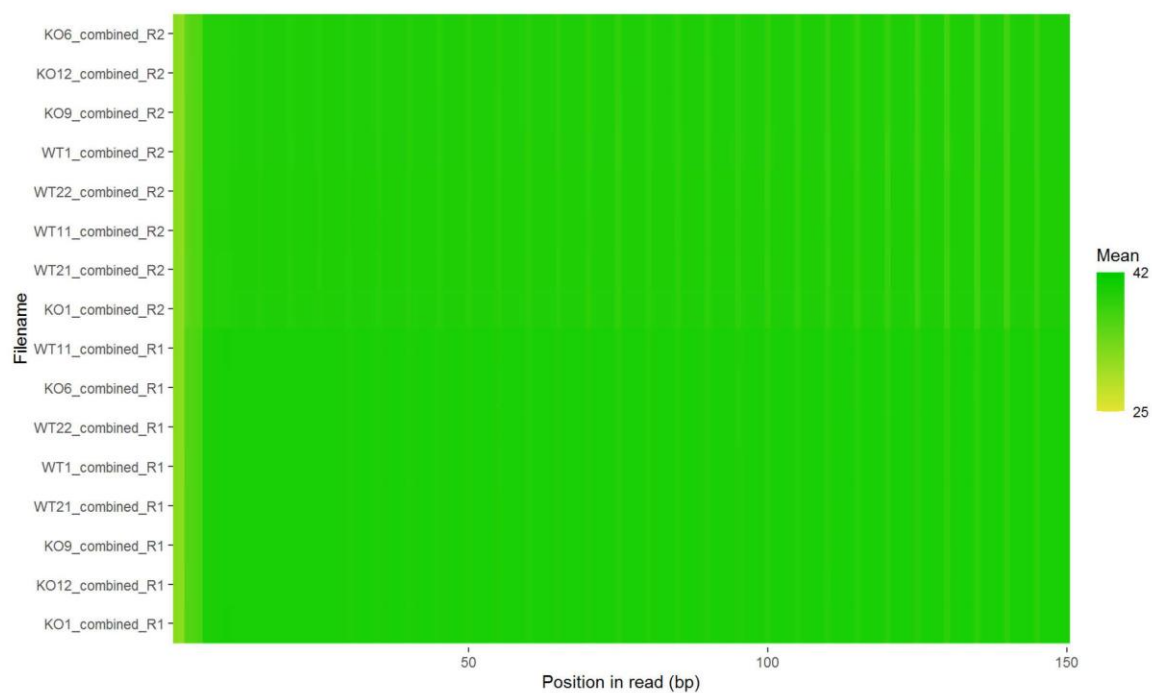
Read totals for each library. Duplicated reads are conventionally a high overestimate at this point.

## FastQC Summary



Summary of FastQC flags for each parameter

## Per Base Sequence Quality



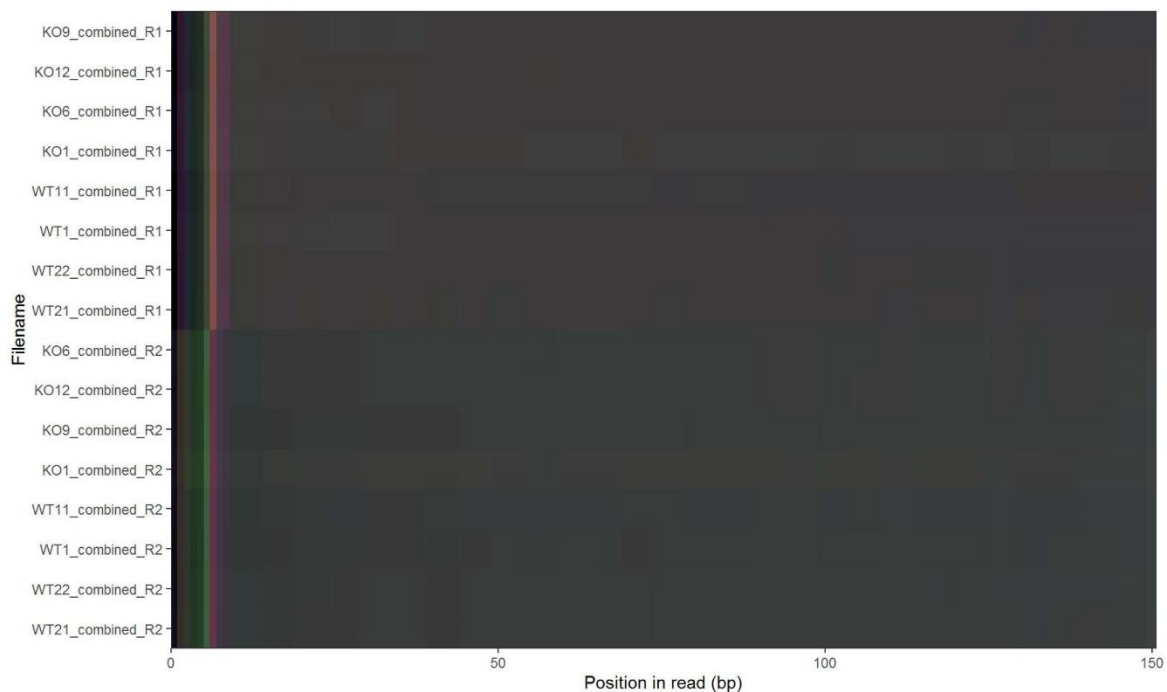
Heatmap showing mean base qualities for each library

## Per Sequence Quality Scores



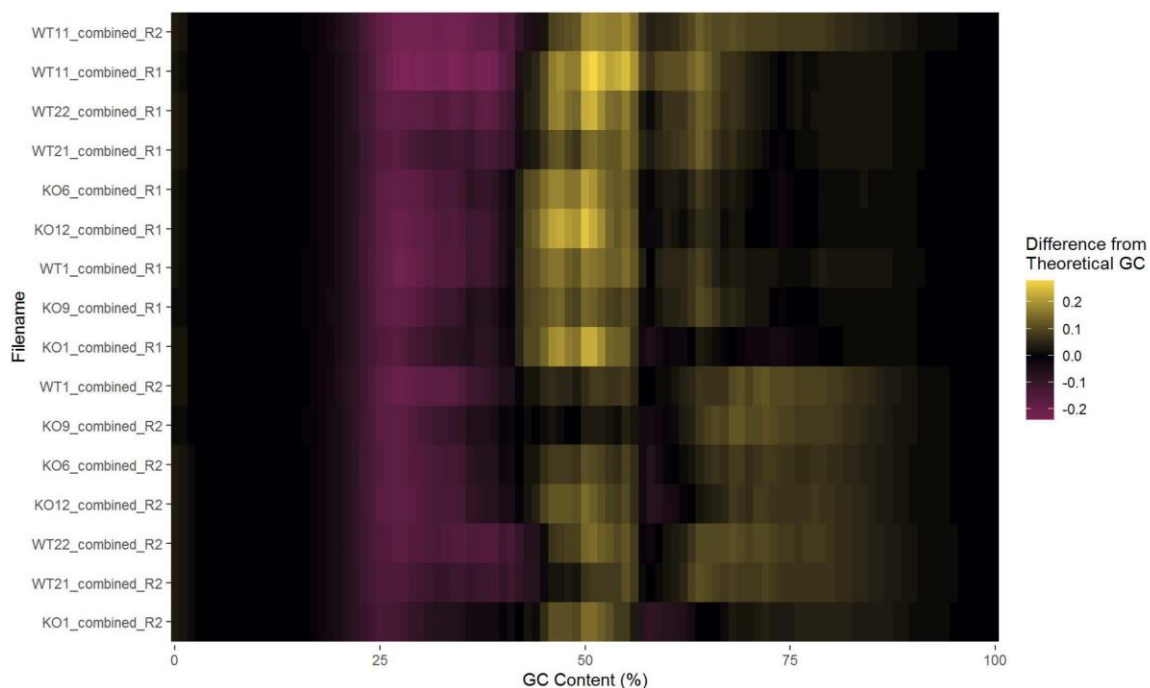
Heatmap showing mean sequence qualities for each library

## Per Base Sequence Content

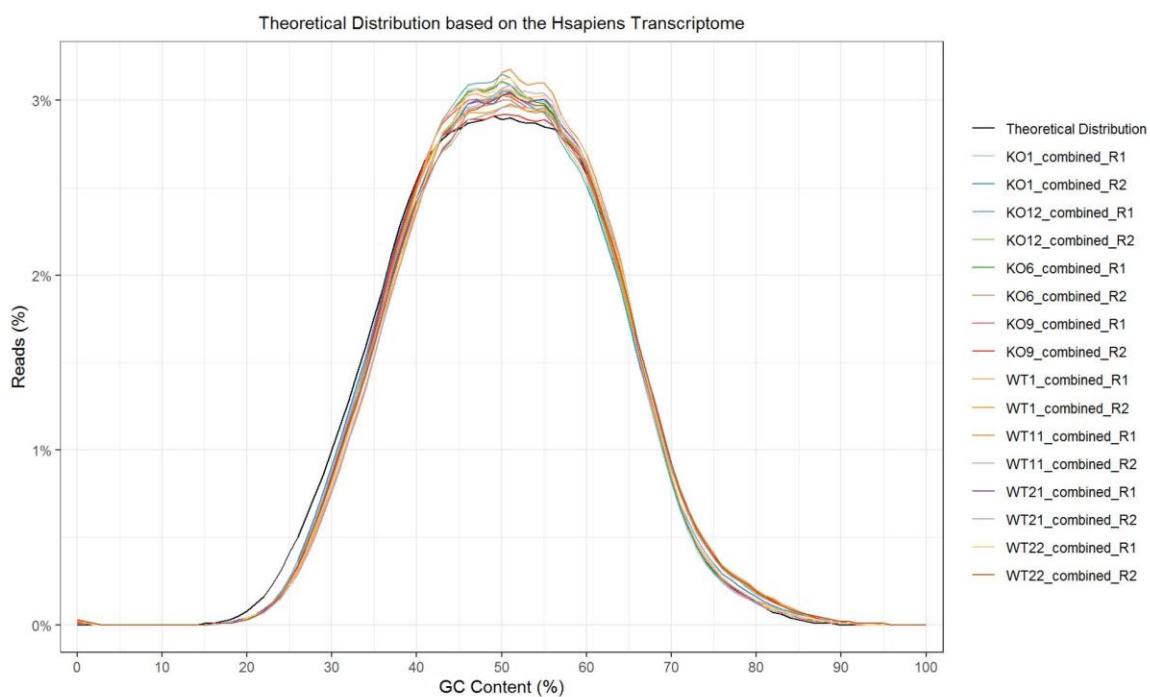


Heatmap of summed base distributions along each read

## Per Sequence GC Content



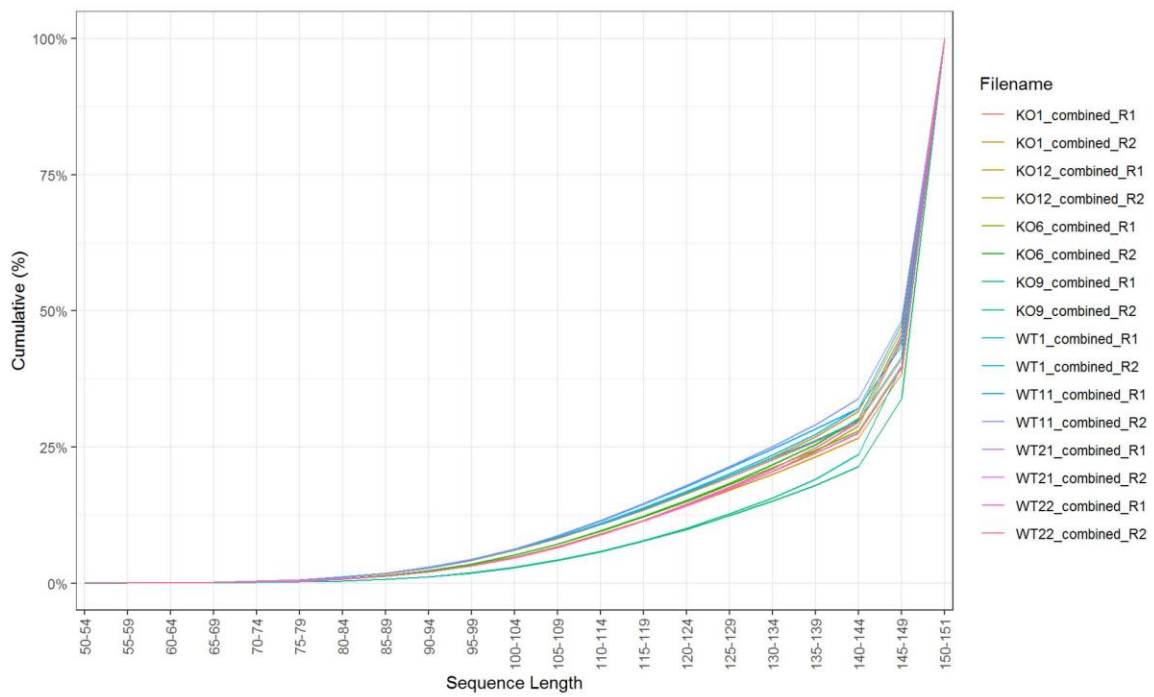
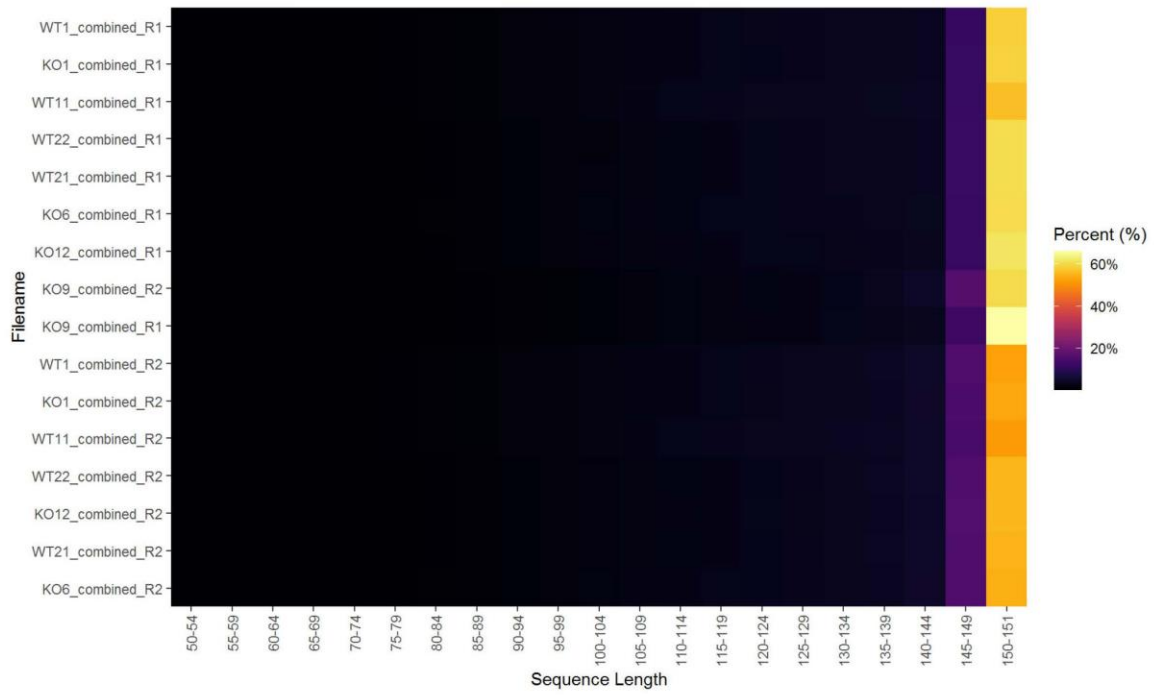
GC Content Heatmap normalised to theoretical GC content in the Hsapiens Transcriptome



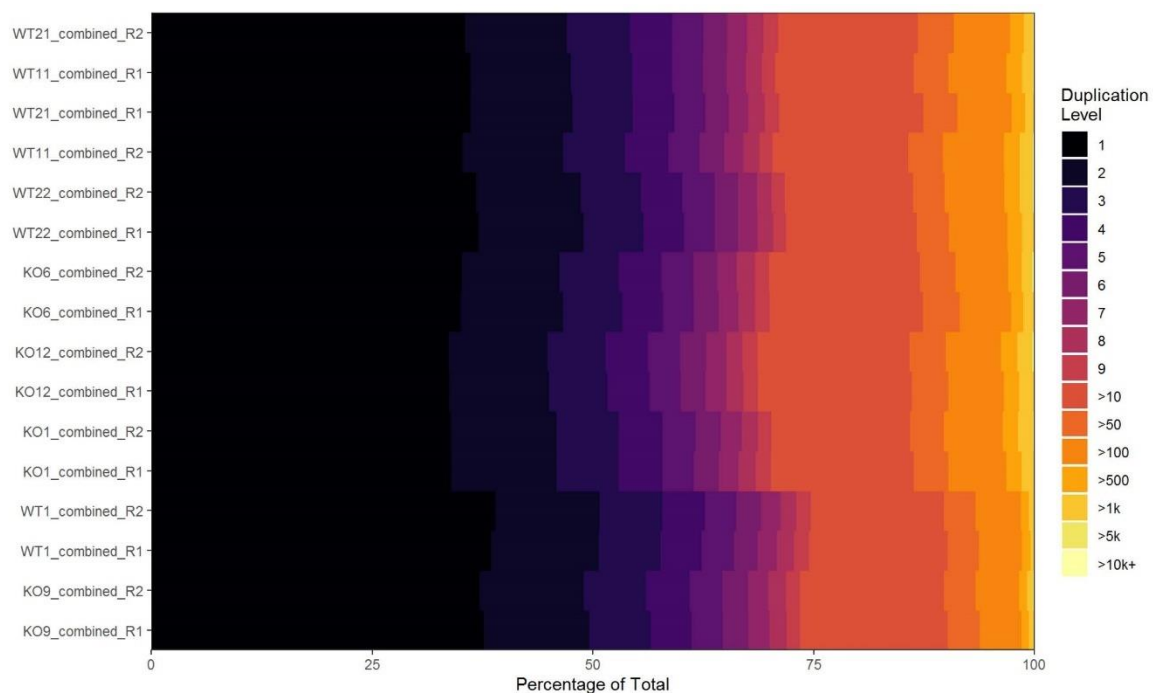
GC Content Distributions for all reads showing theoretical GC content from the Hsapiens Transcriptome

## Sequence Length Distribution

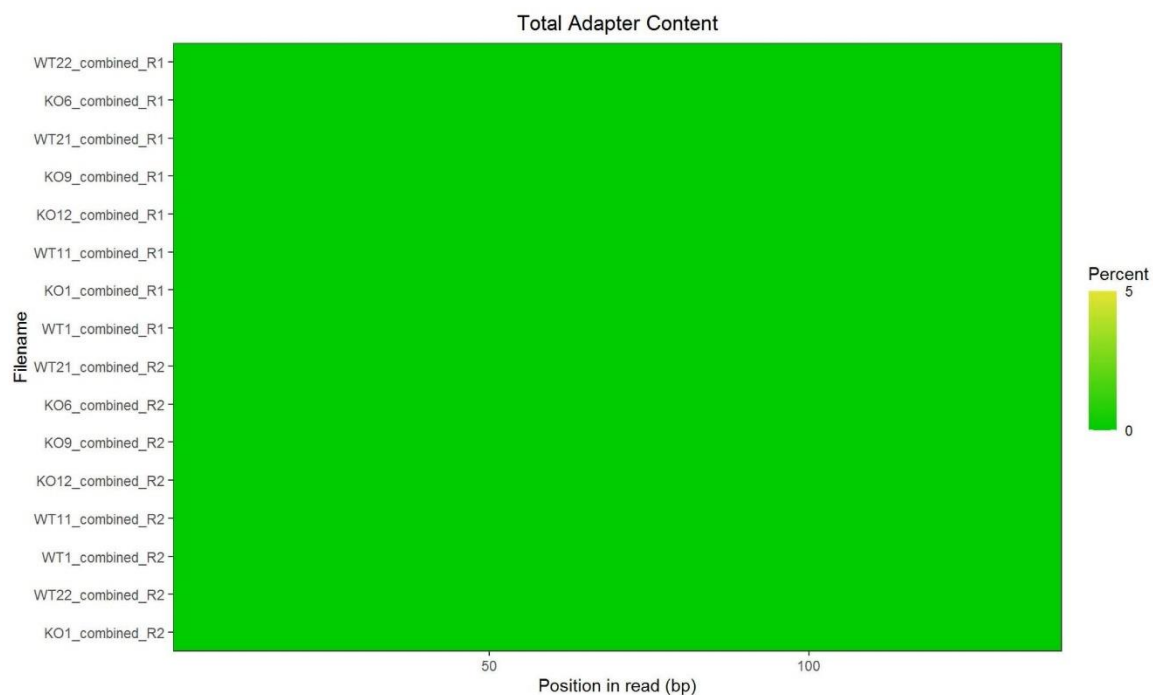




## Sequence Duplication Levels



## Adapter Content



Universal Adapter Content

## Session Information

**R version 4.0.4 (2021-02-15)**

**Platform: x86\_64-w64-mingw32/x64 (64-bit)**

**locale:** LC\_COLLATE=English\_Australia.1252, LC\_CTYPE=English\_Australia.1252, LC\_MONETARY=English\_Australia.1252, LC\_NUMERIC=C and LC\_TIME=English\_Australia.1252

**attached base packages:** parallel, stats, graphics, grDevices, utils, datasets, methods and base

**other attached packages:** pander(v.0.6.3), DT(v.0.17), scales(v.1.1.1), readr(v.1.4.0), dplyr(v.1.0.5), stringr(v.1.4.0), magrittr(v.2.0.1), ngsReports(v.1.6.1), tibble(v.3.1.0), ggplot2(v.3.3.3) and BiocGenerics(v.0.36.0)

**loaded via a namespace (and not attached):** bitops(v.1.0-6), matrixStats(v.0.58.0), lubridate(v.1.7.10), RColorBrewer(v.1.1-2), httr(v.1.4.2), GenomeInfoDb(v.1.26.4), tools(v.4.0.4), bslib(v.0.2.4), utf8(v.1.1.4), R6(v.2.5.0), DBI(v.1.1.1), lazyeval(v.0.2.2), colorspace(v.2.0-0), withr(v.2.4.1), tidyselect(v.1.1.0), compiler(v.4.0.4), Biobase(v.2.50.0), flashClust(v.1.01-2), DelayedArray(v.0.16.2), plotly(v.4.9.3), gg dendro(v.0.1.22), labeling(v.0.4.2), sass(v.0.3.1), digest(v.0.6.27), Rsamtools(v.2.6.0), markdown(v.2.7), XVector(v.0.30.0), jpeg(v.0.1-8.1), pkgconfig(v.2.0.3), htmltools(v.0.5.1.1), MatrixGenerics(v.1.2.1), highr(v.0.8), FactoMineR(v.2.4), htmlwidgets(v.1.5.3), rlang(v.0.4.10), farver(v.2.1.0), jquerylib(v.0.1.3), generics(v.0.1.0), zoo(v.1.8-9), hwriter(v.1.3.2), jsonlite(v.1.7.2), crosstalk(v.1.1.1), BiocParallel(v.1.24.1), RCurl(v.1.98-1.2), GenomeInfoDbData(v.1.2.4), leaps(v.3.1), Matrix(v.1.3-2), Rcpp(v.1.0.6), munsell(v.0.5.0), S4Vectors(v.0.28.1), fansi(v.0.4.2), lifecycle(v.1.0.0), scatterplot3d(v.0.3-41), stringi(v.1.5.3), yaml(v.2.2.1), MASS(v.7.3-53), SummarizedExperiment(v.1.20.0), zlibbioc(v.1.36.0), plyr(v.1.8.6), grid(v.4.0.4), ggrepel(v.0.9.1), forcats(v.0.5.1), crayon(v.1.4.1), lattice(v.0.20-41), Biostrings(v.2.58.0), hms(v.1.0.0), knitr(v.1.31), pillar(v.1.5.1), GenomicRanges(v.1.42.0), reshape2(v.1.4.4), stats4(v.4.0.4), glue(v.1.4.2), evaluate(v.0.14), ShortRead(v.1.48.0), latticeExtra(v.0.6-29), data.table(v.1.14.0), vctrs(v.0.3.6), png(v.0.1-7), gtable(v.0.3.0), purrr(v.0.3.4), tidyr(v.1.1.3), assertthat(v.0.2.1), xfun(v.0.22), viridisLite(v.0.3.0), GenomicAlignments(v.1.26.0), IRanges(v.2.24.1), cluster(v.2.1.0) and ellipsis(v.0.3.1)

## 7.2.2 ngsReports: ZEB2 Knockout in Th1 CM & EM Pools

Vincent Wong

- FastQC Summary
  - Read Totals
  - FastQC Summary
  - Per Base Sequence Quality
  - Per Sequence Quality Scores
  - Per Base Sequence Content
  - Per Sequence GC Content
  - Sequence Length Distribution
  - Sequence Duplication Levels
  - Adapter Content
- Session Information

### FastQC Summary

Show  entries

Search:

Summary statistics for all libraries

Filename	Total	Flagged As Poor Quality	Sequence Length	%GC	File Type	Encoding
08Th1CM- KO1_combined_R1.fastq.gz	26,758,806	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1CM- KO1_combined_R2.fastq.gz	26,758,806	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1CM- KO2_combined_R1.fastq.gz	23,263,928	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1CM- KO2_combined_R2.fastq.gz	23,263,928	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1CM- WT_combined_R1.fastq.gz	28,833,148	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1CM- WT_combined_R2.fastq.gz	28,833,148	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
08Th1EM- KO1_combined_R1.fastq.gz	22,984,938	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1EM- KO1_combined_R2.fastq.gz	22,984,938	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
08Th1EM- KO2_combined_R1.fastq.gz	27,579,352	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
08Th1EM- KO2_combined_R2.fastq.gz	27,579,352	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9

Filename	Total	Flagged As Poor Quality	Sequence Length	%GC	File Type	Encoding
08Th1EM- WT_combined_R1.fastq.gz	31,236,273	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
08Th1EM- WT_combined_R2.fastq.gz	31,236,273	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- KO1_combined_R1.fastq.gz	32,649,030	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- KO1_combined_R2.fastq.gz	32,649,030	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- KO2_combined_R1.fastq.gz	26,404,445	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- KO2_combined_R2.fastq.gz	26,404,445	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- WT_combined_R1.fastq.gz	25,592,039	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1CM- WT_combined_R2.fastq.gz	25,592,039	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- KO1_combined_R1.fastq.gz	34,724,667	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- KO1_combined_R2.fastq.gz	34,724,667	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- KO2_combined_R1.fastq.gz	33,191,375	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- KO2_combined_R2.fastq.gz	33,191,375	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- WT_combined_R1.fastq.gz	32,392,754	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
10Th1EM- WT_combined_R2.fastq.gz	32,392,754	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
16Th1CM- KO_combined_R1.fastq.gz	32,789,398	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1CM- KO_combined_R2.fastq.gz	32,789,398	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1CM- WT_combined_R1.fastq.gz	27,914,468	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1CM- WT_combined_R2.fastq.gz	27,914,468	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9

Filename	Total	Flagged As Poor Quality	Sequence Length	%GC	File Type	Encoding
16Th1EM-KO_combined_R1.fastq.gz	32,396,094	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1EM-KO_combined_R2.fastq.gz	32,396,094	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1EM-WT_combined_R1.fastq.gz	35,623,974	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
16Th1EM-WT_combined_R2.fastq.gz	35,623,974	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
18Th1EM-KO_combined_R1.fastq.gz	36,730,551	0	50-150	49%	Conventional base calls	Sanger / Illumina 1.9
18Th1EM-KO_combined_R2.fastq.gz	36,730,551	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
18Th1EM-WT_combined_R1.fastq.gz	31,084,326	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9
18Th1EM-WT_combined_R2.fastq.gz	31,084,326	0	50-150	50%	Conventional base calls	Sanger / Illumina 1.9

Showing 1 to 36 of 36 entries

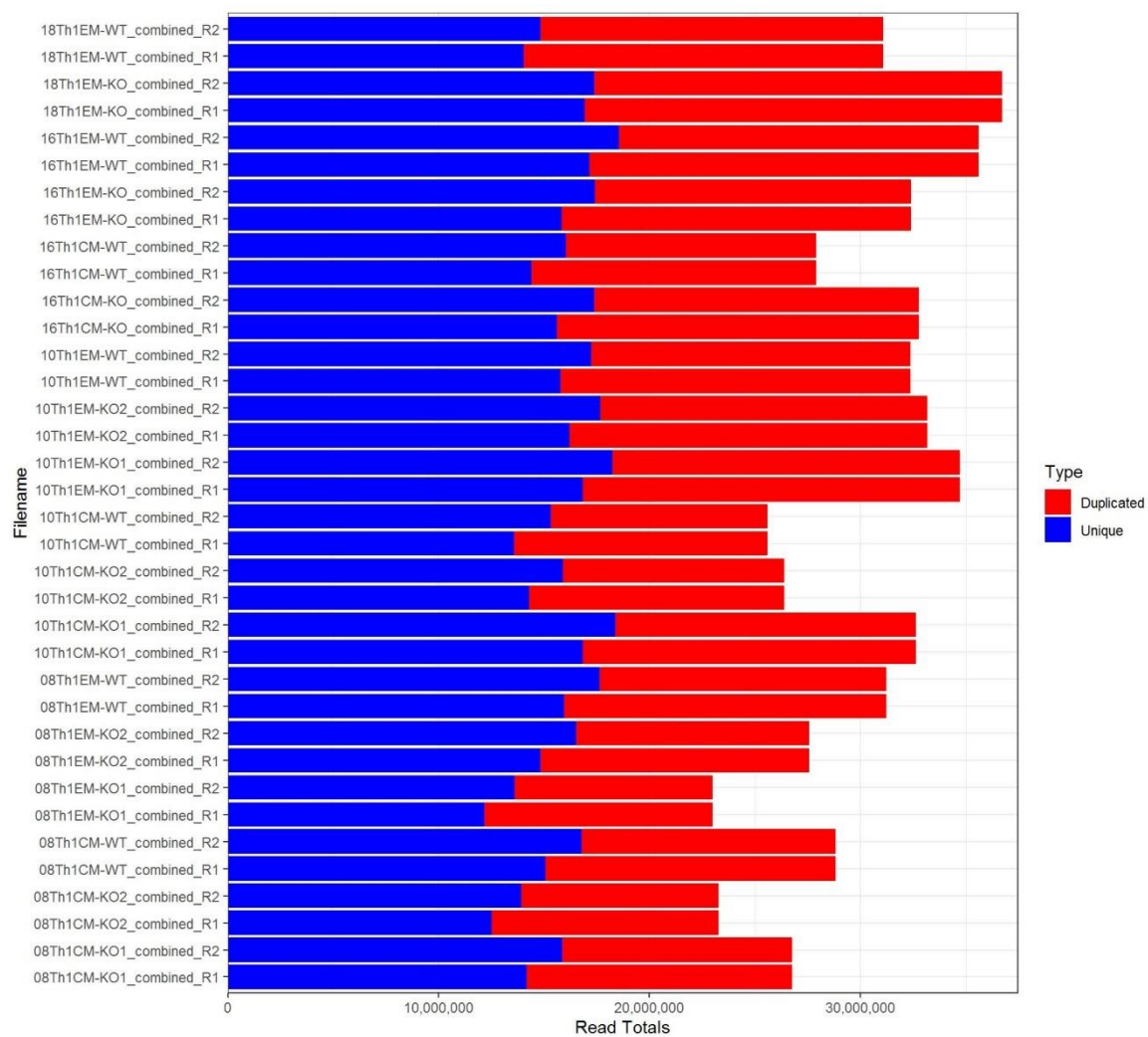
[Previous](#)

1

[Next](#)

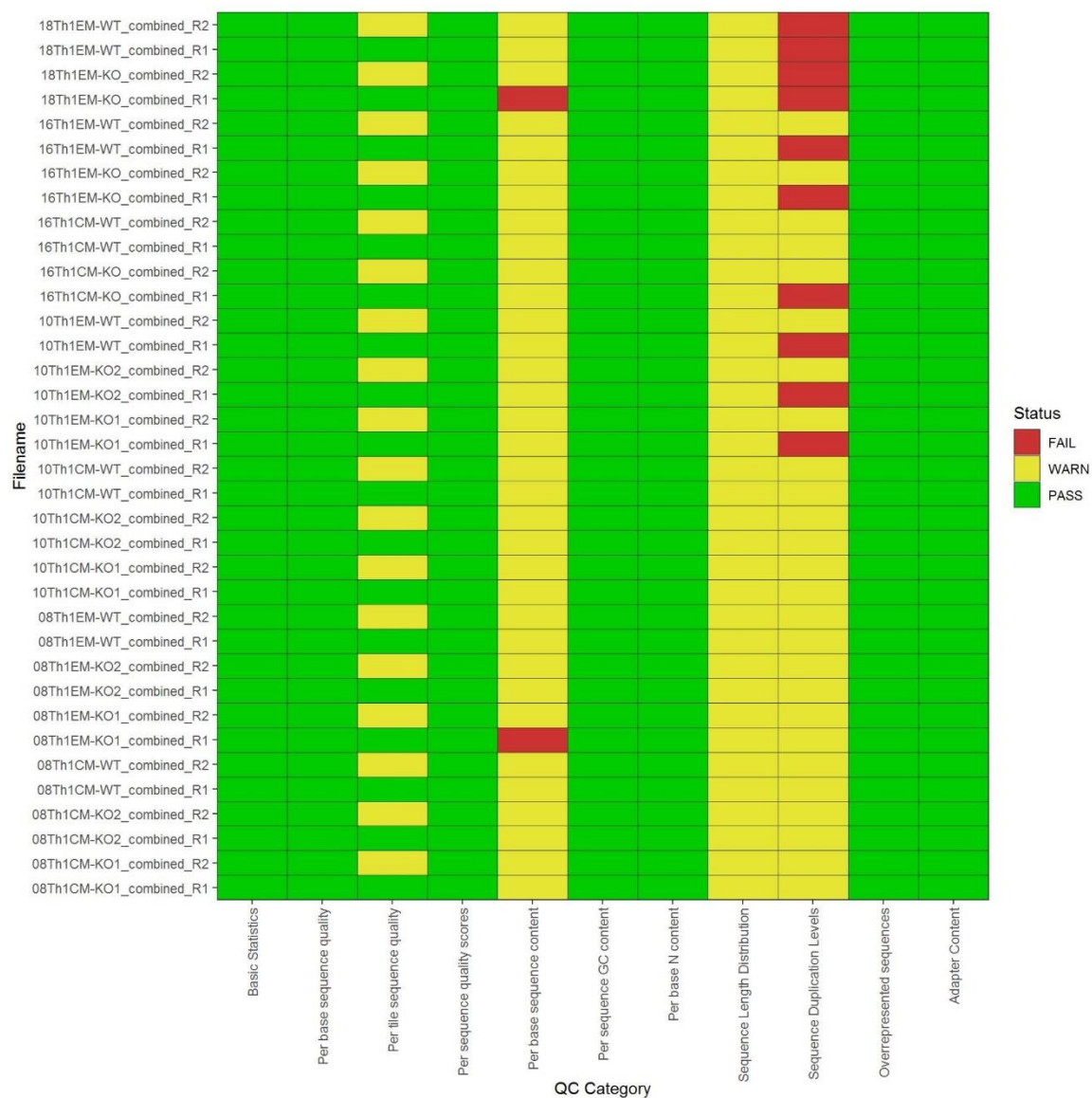
## Read Totals

Library Sizes ranged between 22,984,938 and 36,730,551 reads.



Read totals for each library. Duplicated reads are conventionally a high overestimate at this point.

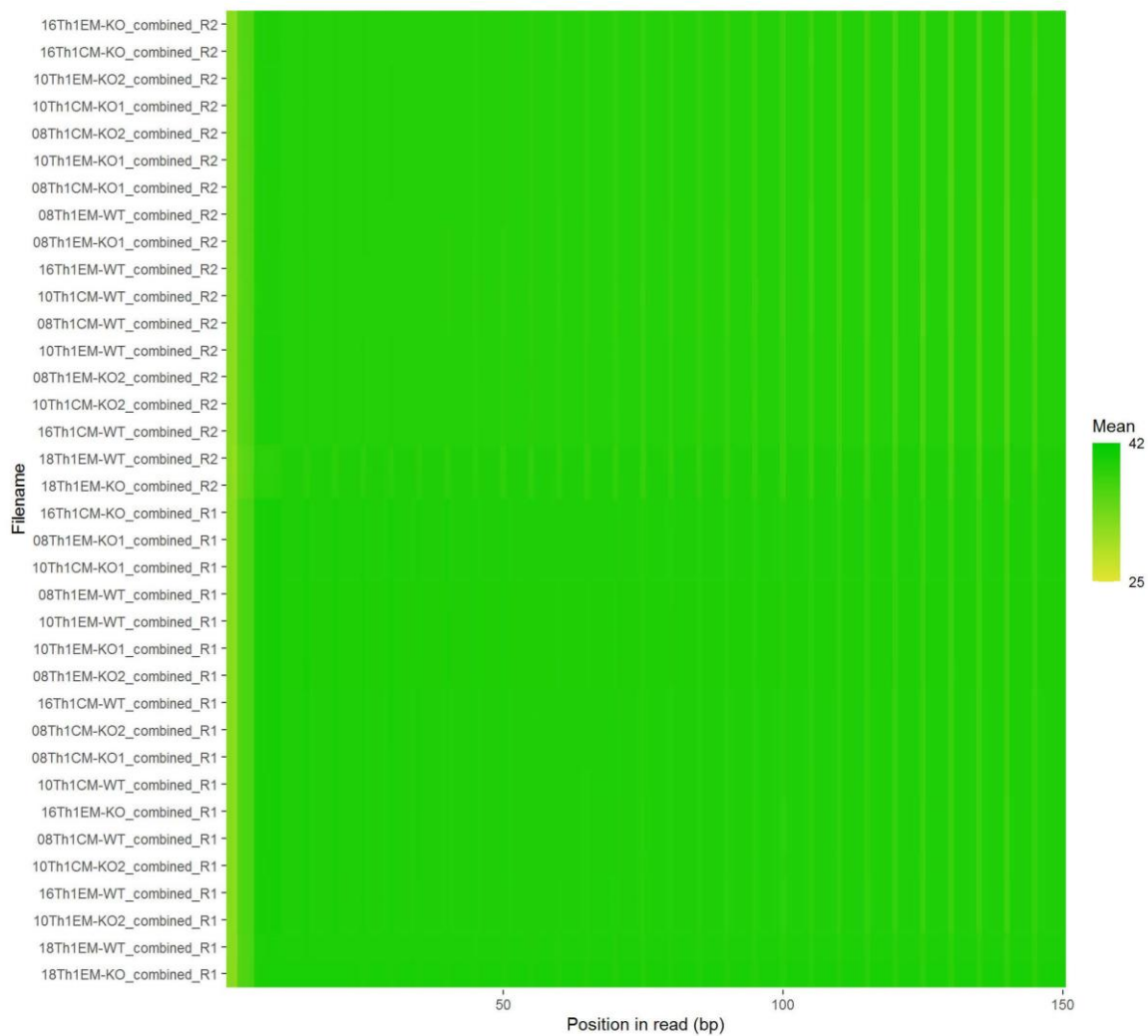
## FastQC Summary



Summary of FastQC flags for each parameter

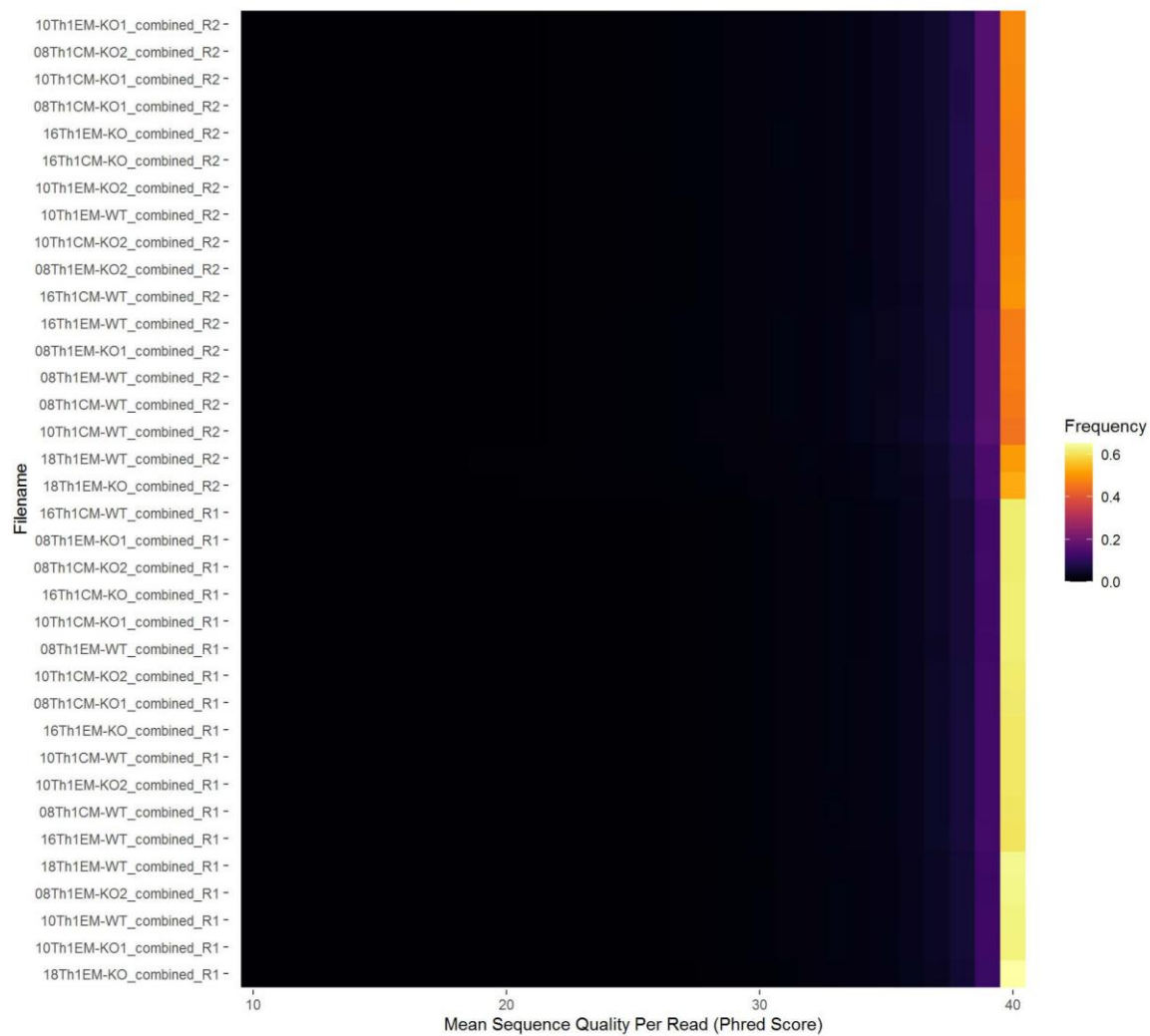
## Per Base Sequence Quality





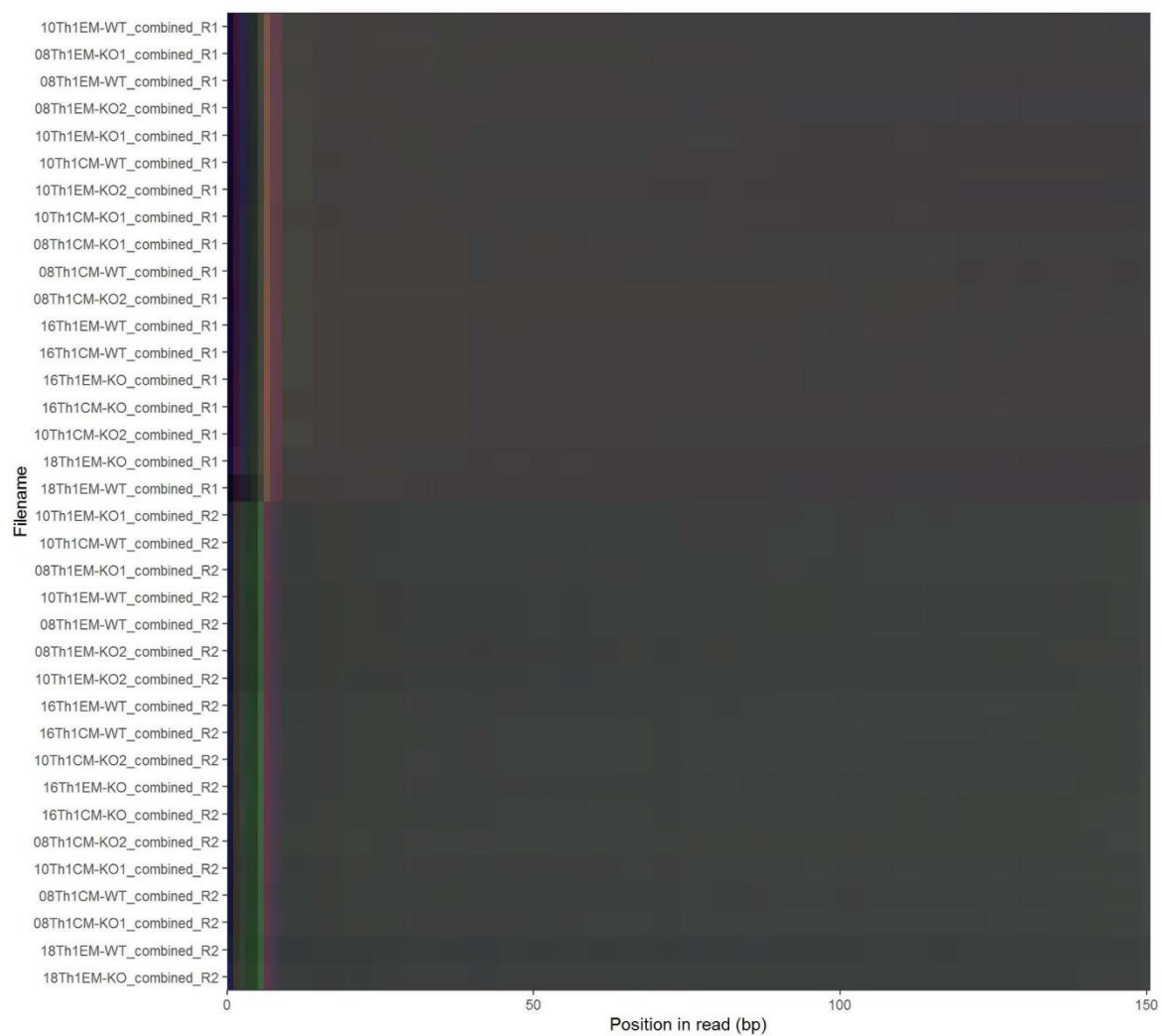
Heatmap showing mean base qualities for each library

## Per Sequence Quality Scores



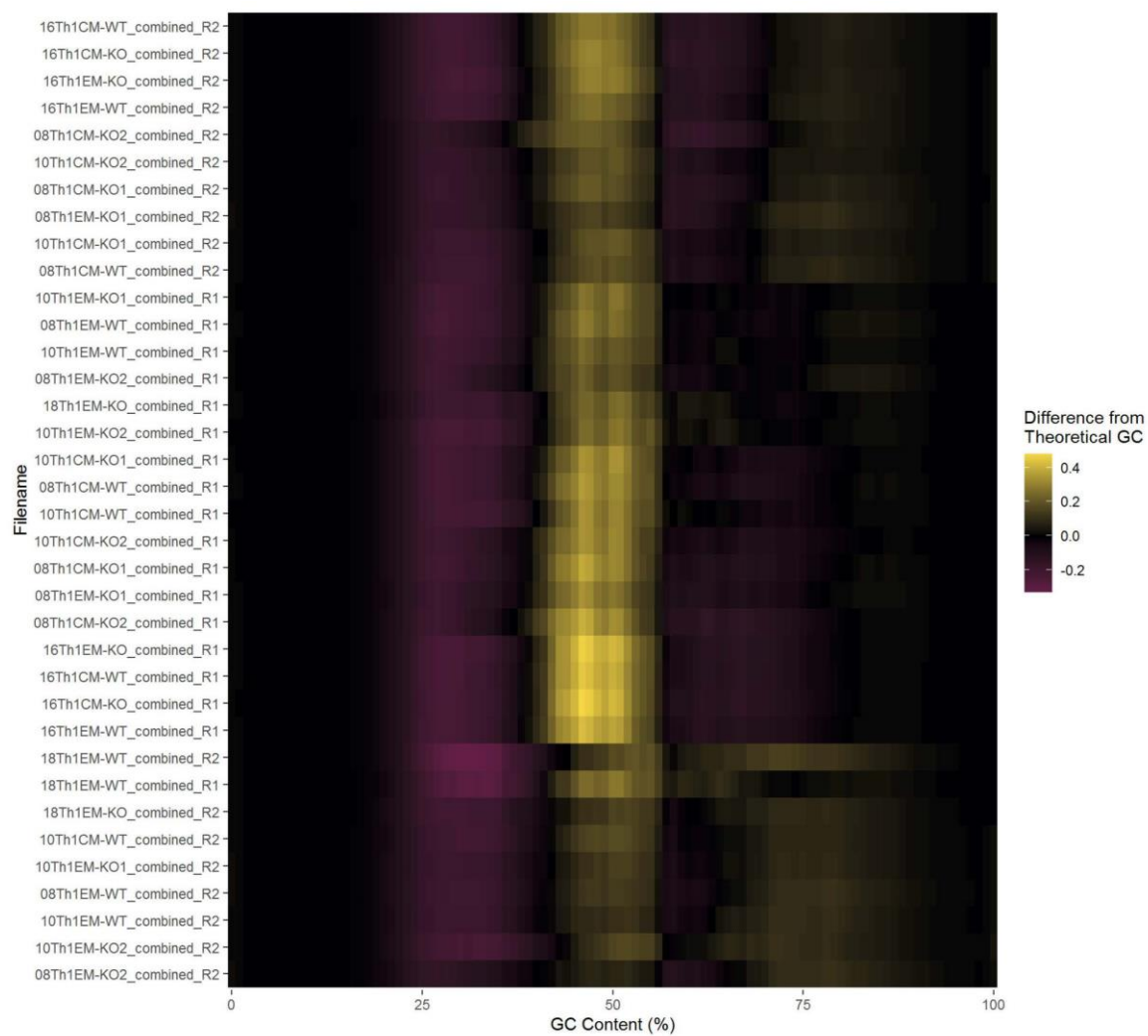
Heatmap showing mean sequence qualities for each library

## Per Base Sequence Content

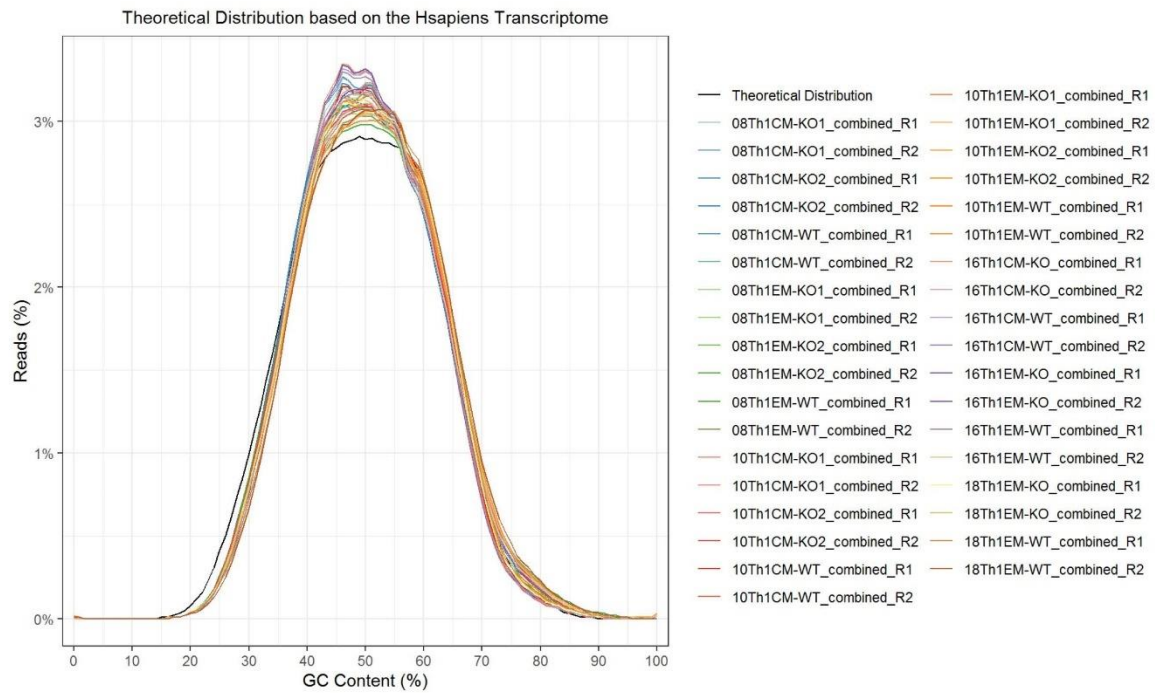


Heatmap of summed base distributions along each read

## Per Sequence GC Content

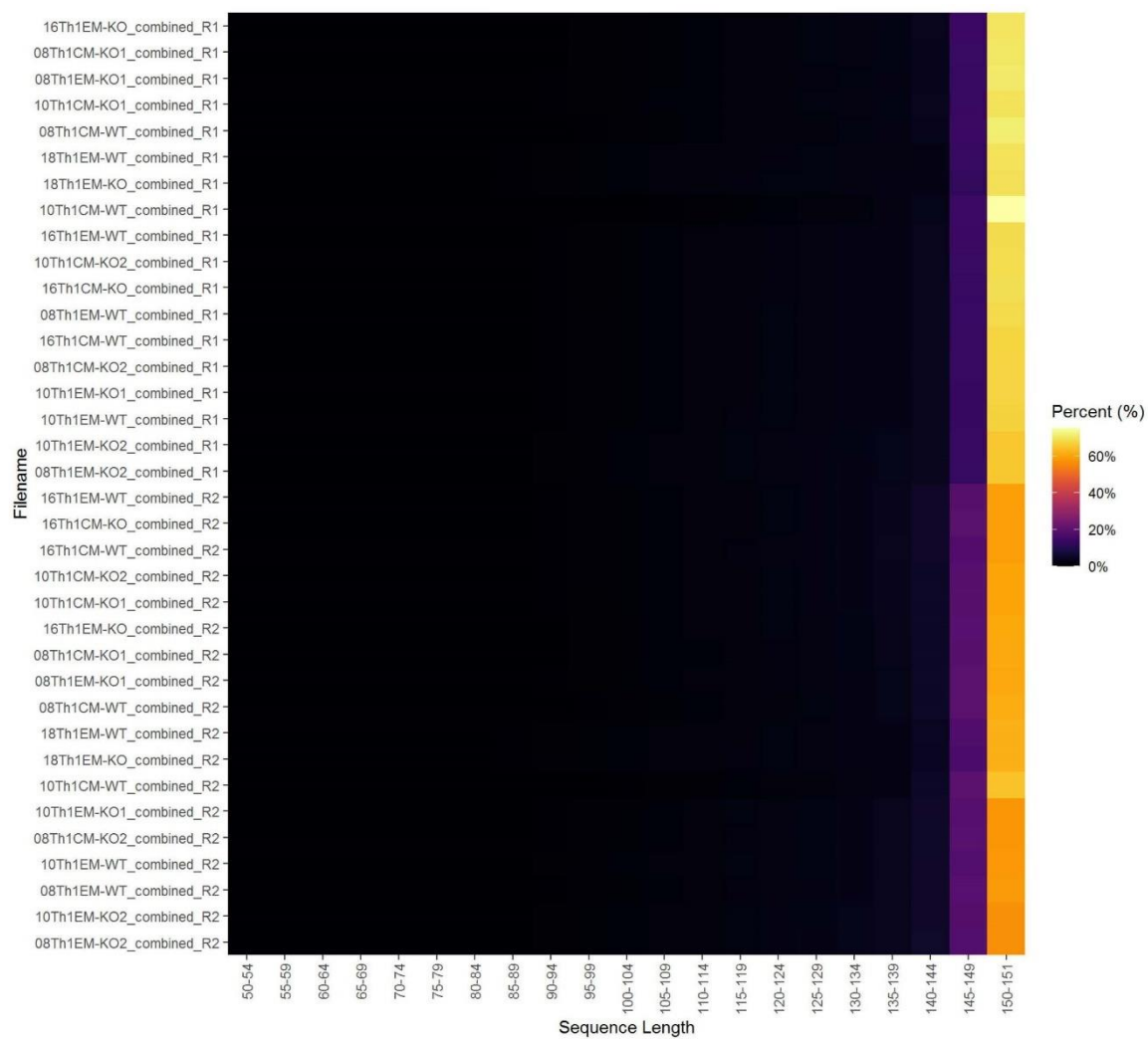


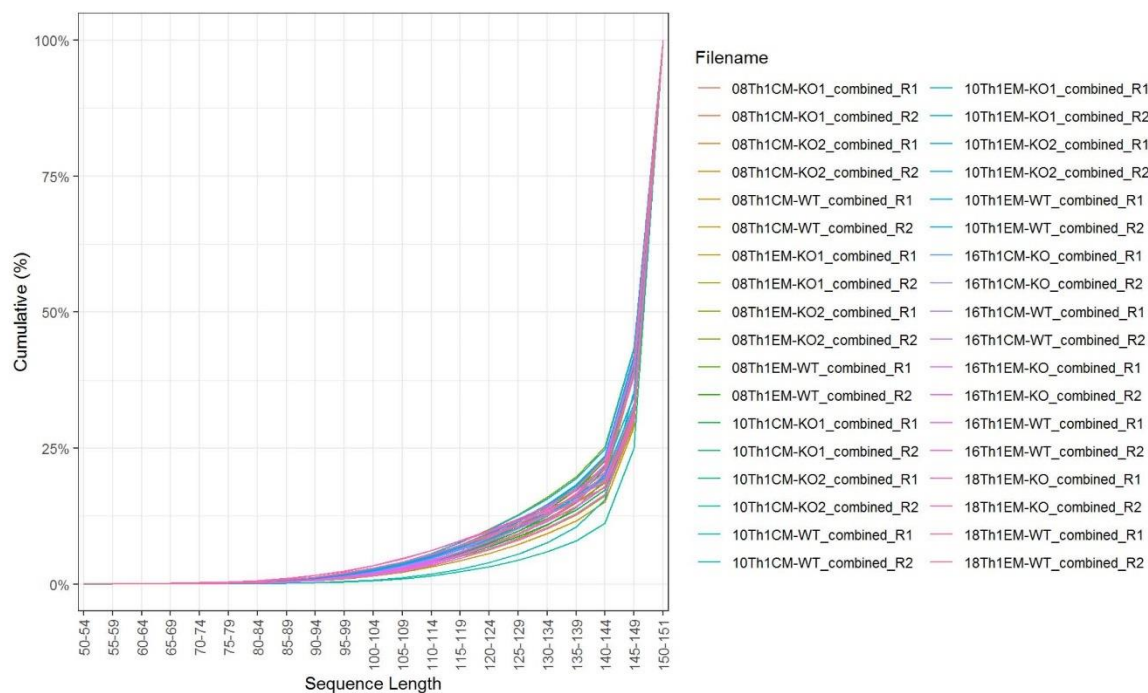
GC Content Heatmap normalised to theoretical GC content in the Hsapiens Transcriptome



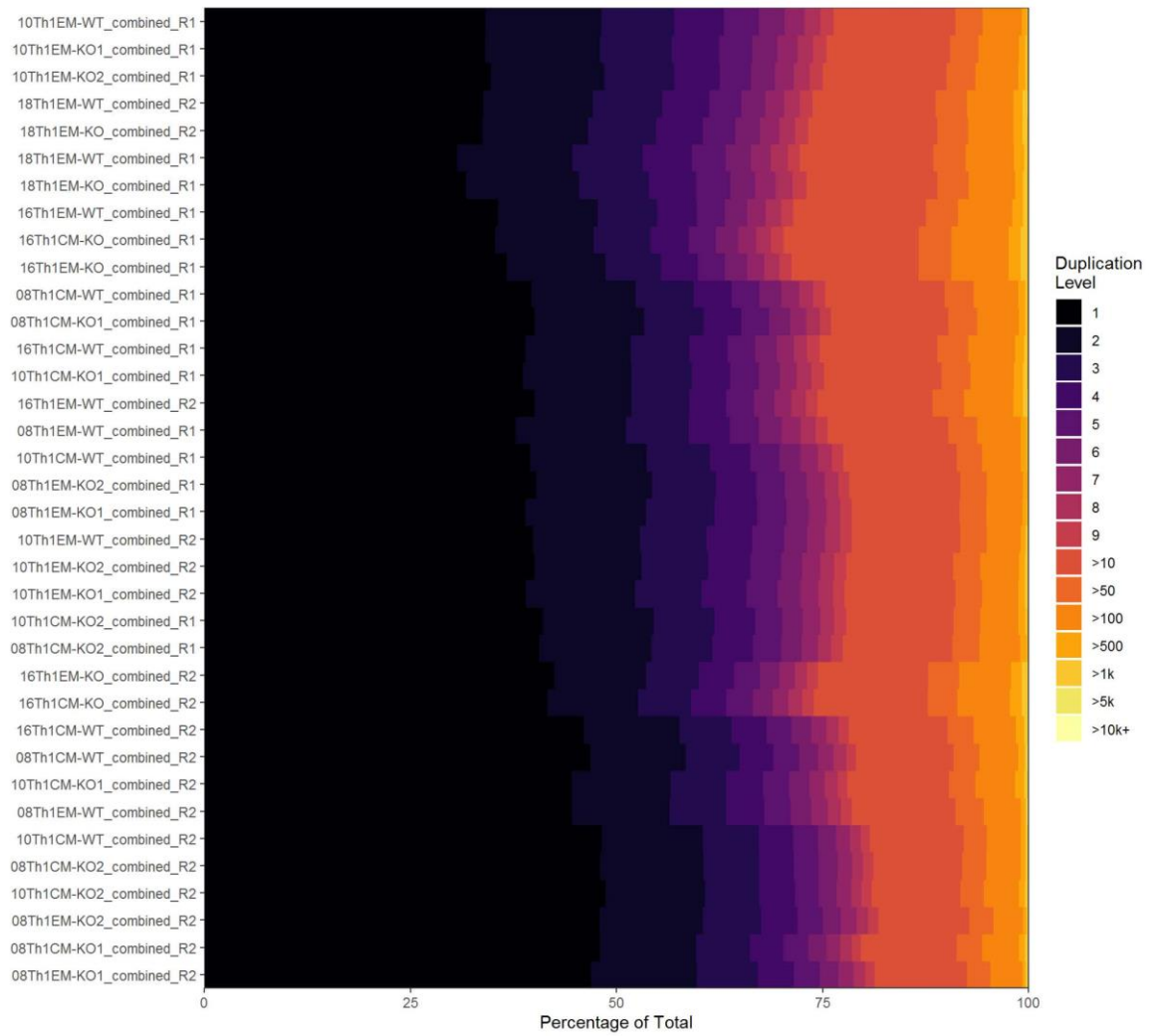
GC Content Distributions for all reads showing theoretical GC content from the Hsapiens Transcriptome

## Sequence Length Distribution



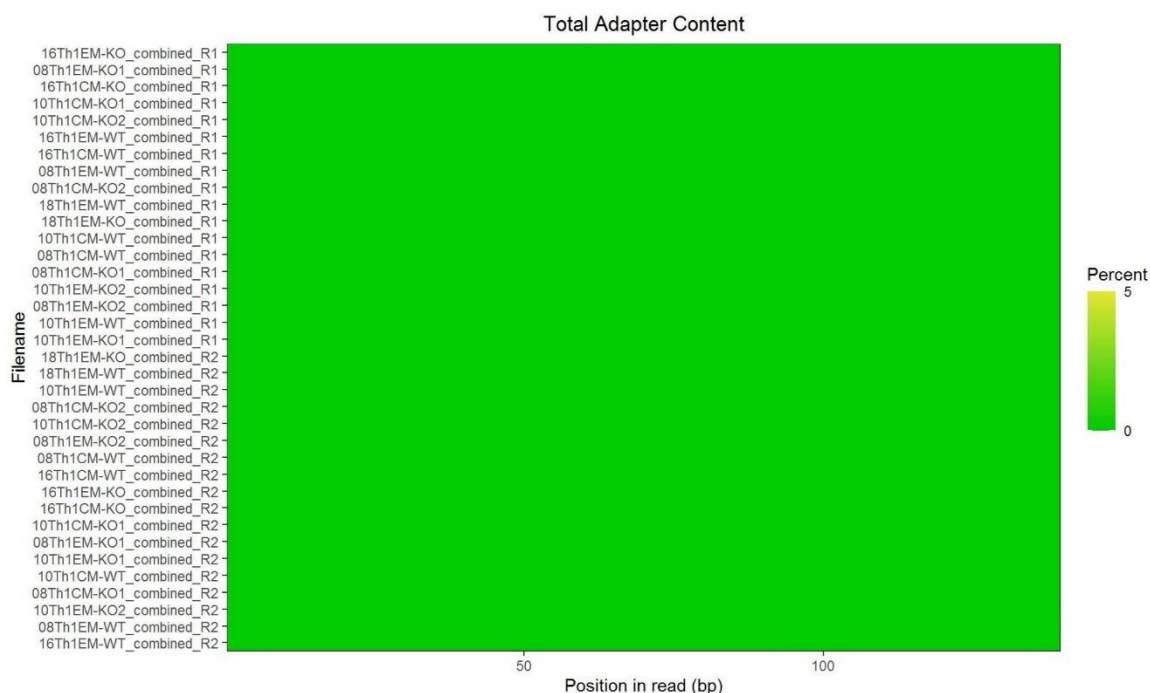


## Sequence Duplication Levels



## Adapter Content





Universal Adapter Content

## Session Information

**R version 4.0.4 (2021-02-15)**

**Platform:** x86\_64-w64-mingw32/x64 (64-bit)

**locale:** LC\_COLLATE=English\_Australia.1252, LC\_CTYPE=English\_Australia.1252, LC\_MONETARY=English\_Australia.1252, LC\_NUMERIC=C and LC\_TIME=English\_Australia.1252

**attached base packages:** *parallel*, *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

**other attached packages:** *pander*(v.0.6.3), *DT*(v.0.17), *scales*(v.1.1.1), *readr*(v.1.4.0), *dplyr*(v.1.0.5), *stringr*(v.1.4.0), *magrittr*(v.2.0.1), *ngsReports*(v.1.6.1), *tibble*(v.3.1.0), *ggplot2*(v.3.3.3) and *BiocGenerics*(v.0.36.0)

**loaded via a namespace (and not attached):** *bitops*(v.1.0-6), *matrixStats*(v.0.58.0), *lubridate*(v.1.7.10), *RColorBrewer*(v.1.1-2), *httr*(v.1.4.2), *GenomeInfoDb*(v.1.26.4), *tools*(v.4.0.4), *bslib*(v.0.2.4), *utf8*(v.1.1.4), *R6*(v.2.5.0), *DBI*(v.1.1.1), *lazyeval*(v.0.2.2), *colorspace*(v.2.0-0), *withr*(v.2.4.1), *tidyselect*(v.1.1.0), *compiler*(v.4.0.4), *Biobase*(v.2.50.0), *flashClust*(v.1.01-2), *DelayedArray*(v.0.16.2), *plotly*(v.4.9.3), *ggdendro*(v.0.1.22), *labeling*(v.0.4.2), *sass*(v.0.3.1), *digest*(v.0.6.27), *Rsamtools*(v.2.6.0), *rmarkdown*(v.2.7), *XVector*(v.0.30.0), *jpeg*(v.0.1-8.1), *pkgconfig*(v.2.0.3), *htmltools*(v.0.5.1.1), *MatrixGenerics*(v.1.2.1), *highr*(v.0.8), *FactoMineR*(v.2.4), *htmlwidgets*(v.1.5.3), *rlang*(v.0.4.10), *farver*(v.2.1.0), *jquerylib*(v.0.1.3), *generics*(v.0.1.0), *zoo*(v.1.8-9), *hwriter*(v.1.3.2), *jsonlite*(v.1.7.2), *crosstalk*(v.1.1.1), *BiocParallel*(v.1.24.1), *RCurl*(v.1.98-1.2), *GenomeInfoDbData*(v.1.2.4), *leaps*(v.3.1), *Matrix*(v.1.3-2), *Rcpp*(v.1.0.6), *munsell*(v.0.5.0), *S4Vectors*(v.0.28.1), *fansi*(v.0.4.2), *lifecycle*(v.1.0.0), *scatterplot3d*(v.0.3-41), *stringi*(v.1.5.3), *yaml*(v.2.2.1), *MASS*(v.7.3-53), *SummarizedExperiment*(v.1.20.0), *zlibbioc*(v.1.36.0), *plyr*(v.1.8.6), *grid*(v.4.0.4), *ggrepel*(v.0.9.1), *forcats*(v.0.5.1), *crayon*(v.1.4.1), *lattice*(v.0.20-41), *Biostrings*(v.2.58.0), *hms*(v.1.0.0), *knitr*(v.1.31), *pillar*(v.1.5.1), *GenomicRanges*(v.1.42.0), *reshape2*(v.1.4.4), *stats4*(v.4.0.4), *glue*(v.1.4.2), *evaluate*(v.0.14), *ShortRead*(v.1.48.0), *latticeExtra*(v.0.6-29), *data.table*(v.1.14.0), *vctrs*(v.0.3.6), *png*(v.0.1-7), *gtable*(v.0.3.0), *purrr*(v.0.3.4), *tidyr*(v.1.1.3), *assertthat*(v.0.2.1), *xfun*(v.0.22), *viridisLite*(v.0.3.0), *GenomicAlignments*(v.1.26.0), *IRanges*(v.2.24.1), *cluster*(v.2.1.0) and *ellipsis*(v.0.3.1)

## 7.3 RNA-seq bash Scripts

### 7.3.1 Bash script for ZEB2 knockout in Th1 EM clones

#### [Pipeline script](#)

```

1. #!/bin/bash
2. #SBATCH -p batch
3. #SBATCH -N 1
4. #SBATCH -n 16
5. #SBATCH --time=24:00:00
6. #SBATCH --mem=48GB
7. #SBATCH --mail-type=END
8. #SBATCH --mail-type=FAIL
9. #SBATCH --mail-user=soonwei.wong@adelaide.edu.au
10.
11. ## Note that the FastQC reports on the raw data were previously
    generated
12. ## for a preliminary meeting
13.
14. ## Cores
15. CORES=16
16.
17. ## Modules
18. module load FastQC/0.11.7
19. module load STAR/2.7.0d-foss-2016b
20. module load SAMtools/1.10-foss-2016b
21. module load AdapterRemoval/2.2.1-foss-2016b
22. module load GCC/9.1.0-2.32
23. module load Subread/1.5.2-foss-2016b
24.
25. ## Genomic Data Files
26. REFS=/data/biorefs/reference_genomes/ensembl-release-
    98/homo_sapiens
27. GTF=/fast/users/a1656791/Th1_ZEB2_KO/070619/steve_script/Homo_sapie
    ns.GRCh38.98.chr.gtf
28.
29. ## Directories
30. PROJROOT=/fast/users/a1656791/Th1_ZEB2_KO/271118/steve_script
31.
32. ## Directories for Initial FastQC
33. RAWDATA=$PROJROOT/0_rawData
34. # mkdir -p $RAWDATA/FastQC
35.
36. ## Setup for Trimmed data
37. TRIMDATA=$PROJROOT/1_trimmedData
38. mkdir -p $TRIMDATA/fastq
39. mkdir -p $TRIMDATA/FastQC
40. mkdir -p $TRIMDATA/logs
41.
42. ## Setup for genome alignment
43. ALIGNDATA=$PROJROOT/2_alignedData
44. mkdir -p $ALIGNDATA/logs
45. mkdir -p $ALIGNDATA/bams
46. mkdir -p $ALIGNDATA/FastQC
47. mkdir -p $ALIGNDATA/featureCounts
48.
49. ##-----##

```

```

50. ## Concatenate on the raw data (if required)
51. ##-----##
52.
53. ## Setup for Lanes Data
54. LANE1=$PROJROOT/0_rawData/fastq/Lane1
55. LANE2=$PROJROOT/0_rawData/fastq/Lane2
56.
57. ## Setup for Merged Data
58. mkdir -p $PROJROOT/0_rawData/fastq/merged
59. MERGED=$PROJROOT/0_rawData/fastq/merged
60.
61. #for i in $LANE1/*fastq.gz
62.     do
63.
64.         # Now create the output filenames
65.         out=$MERGED/$(basename $i)
66.
67.         # Next create lanes
68.         L1=$LANE1/$(basename $i)
69.         L2=$LANE2/$(basename $i)
70.
71.         echo -e "Currently concatenating $(basename $i) across lanes"
72.         #Concatenate
73.         cat $L1 $L2 > $out
74.
75.     done
76.
77. ##-----##
78. ## FastQC on the raw data (not required)
79. ##-----##
80.
81. # fastqc -t $CORES -o $RAWDATA/FastQC --noextract
    $RAWDATA/fastq/*fastq.gz
82.
83. ##-----##
84. ## Trimming the Merged data
85. ##-----##
86.
87. for R1 in ${RAWDATA}/fastq/*R1.fastq.gz
88.     do
89.
90.         echo -e "Currently working on ${R1}"
91.
92.         # Now create the output filenames
93.         out1=${TRIMDATA}/fastq/$(basename $R1)
94.         BNAME=${TRIMDATA}/fastq/$(basename ${R1%_R1.fastq.gz})
95.         echo -e "Output file 1 will be ${out1}"
96.
97.         ## Repeat for R2 reads
98.         R2=${R1%_R1.fastq.gz}_R2.fastq.gz
99.         out2=${out1%_R1.fastq.gz}_R2.fastq.gz
100.
101.        echo -e "Trimming:\t${R1}\n\t${R2}"
102.        # Trim
103.        AdapterRemoval \
104.            --gzip \
105.            --trimns \
106.            --trimqualities \
107.            --minquality 30 \
108.            --minlength 50 \
109.            --threads ${CORES} \

```

```

110.     --basename ${BNAME} \
111.     --output1 ${out1} \
112.     --output2 ${out2} \
113.     --file1 ${R1} \
114.     --file2 ${R2}
115.
116.     done
117.
118. # Move the log files into their own folder
119. mv ${TRIMDATA}/fastq/*settings ${TRIMDATA}/logs
120.
121. # Run FastQC
122. fastqc -t ${CORES} -o ${TRIMDATA}/FastQC --noextract
    ${TRIMDATA}/fastq/*fastq.gz
123.
124.
125.
126. ##-----##
127. ## salmon
128. ##-----##
129.
130. ## Just submit these as independent jobs at this point as we don't
    need to do any
131. ## post alignment QC now
132. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
133.     do
134.         sbatch ${PROJROOT}/bash/salmonSingle.sh ${R1}
135.     done
136.
137.
138.
139. ##-----##
140. ## Aligning trimmed data to the genome
141. ##-----##
142.
143. ## Aligning, filtering and sorting
144. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
145.     do
146.
147.         R2=${R1%_R1.fastq.gz}_R2.fastq.gz
148.         BNAME=$(basename ${R1%_R1.fastq.gz})
149.         echo -e "STAR will align:\t${R1}\n\t${R2}"
150.
151.         STAR \
152.             --runThreadN ${CORES} \
153.             --genomeDir ${REFS}/star \
154.             --readFilesIn ${R1} ${R2} \
155.             --readFilesCommand gunzip -c \
156.             --outFileNamePrefix ${ALIGNDATA}/bams/${BNAME} \
157.             --outSAMtype BAM SortedByCoordinate
158.
159.     done
160.
161. # Move the log files into their own folder
162. mv ${ALIGNDATA}/bams/*out ${ALIGNDATA}/logs
163. mv ${ALIGNDATA}/bams/*tab ${ALIGNDATA}/logs
164.
165. # Fastqc and indexing
166. for BAM in ${ALIGNDATA}/bams/*.bam
167.     do

```

```
168. fastqc -t ${CORES} -f bam_mapped -o ${ALIGNDATA}/FastQC --
    noextract $(Bamias et al.)
169. samtools index $(Bamias et al.)
170. done
171.
172.
173. ##-----##
174. ## featureCounts
175. ##-----##
176.
177. ## Feature Counts - obtaining all sorted bam files
178. sampleList=`find ${ALIGNDATA}/bams -name "*out.bam" | tr '\n' ' '`
179.
180. ## Running featureCounts on the sorted bam files
181. featureCounts -Q 10 \
182.   -p \
183.   -s 2 \
184.   -T ${CORES} \
185.   -a ${GTF} \
186.   -o ${ALIGNDATA}/featureCounts/counts.out ${sampleList}
187.
188. ## Storing the output in a single file
189. cut -f1,7- ${ALIGNDATA}/featureCounts/counts.out | \
190. sed 1d > ${ALIGNDATA}/featureCounts/genes.out
```

### Additional script for Salmon

```
1. #!/bin/bash
2. #SBATCH -p batch
3. #SBATCH -N 1
4. #SBATCH -n 1
5. #SBATCH --time=03:30:00
6. #SBATCH --mem=4GB
7. #SBATCH --mail-type=END
8. #SBATCH --mail-type=FAIL
9. #SBATCH --mail-user=soonwei.wong@adelaide.edu.au
10.
11. ## Cores
12. CORES=16
13.
14. # Load modules
15. module load Anaconda3/2019.03
16. source activate SalmonEnv
17. module load SAMtools/1.10-foss-2016b
18.
19. ## Genomic Data Files
20. IDX=/fast/users/a1656791/Salmon/salmon_index
21.
22. ## Directories
23. PROJROOT=/fast/users/a1656791/Th1_ZEB2_KO/271118/steve_script
24. TRIMDATA=${PROJROOT}/1_trimmedData
25.
26. ## Setup for salmon output
27. ALIGNDATA=${PROJROOT}/4_salmon
28.
29. ## Now organise the input files
30. F1=$1
31. F2=${F1%R1.fastq.gz}R2.fastq.gz
32.
33. ## Organise the output files
34. OUTDIR=${ALIGNDATA}/${basename ${F1%_R1.fastq.gz}}
35. echo -e "Creating ${OUTDIR}"
36. mkdir -p ${OUTDIR}
37.
38. echo -e "Currently aligning:\n\t${F1}\n\t${F2}"
39. echo -e "Output will be written to ${OUTDIR}"
40.
41. ## Check salmon version
42. salmon --version
43.
44. ## Transcript quantification with salmon quant
45. salmon quant \
46.     -i ${IDX} \
47.     -l A \
48.     -1 ${F1} -2 ${F2} \
49.     --useVBOpt \
50.     --seqBias \
51.     --numBootstraps 50 \
52.     --validateMappings \
53.     --threads ${CORES} \
54.     -o ${OUTDIR}
```

## 7.3.2 Bash script for ZEB2 knockout in Th1 EM pools

### Pipeline script

```

1. #!/bin/bash
2. #SBATCH -p batch
3. #SBATCH -N 1
4. #SBATCH -n 16
5. #SBATCH --time=24:00:00
6. #SBATCH --mem=48GB
7. #SBATCH --mail-type=END
8. #SBATCH --mail-type=FAIL
9. #SBATCH --mail-user=soonwei.wong@adelaide.edu.au
10.
11. ## Note that the FastQC reports on the raw data were previously
    generated
12. ## for a preliminary meeting
13.
14. ## Cores
15. CORES=16
16.
17. ## Modules
18. module load FastQC/0.11.7
19. module load STAR/2.7.0d-foss-2016b
20. module load SAMtools/1.10-foss-2016b
21. module load AdapterRemoval/2.2.1-foss-2016b
22. module load GCC/9.1.0-2.32
23. module load Subread/1.5.2-foss-2016b
24.
25. ## Genomic Data Files
26. REFS=/data/biorefs/reference_genomes/ensembl-release-
    98/homo_sapiens
27. GTF=/fast/users/a1656791/Th1_ZEB2_KO/070619/steve_script/Homo_sapie
    ns.GRCh38.98.chr.gtf
28.
29. ## Directories
30. PROJROOT=/fast/users/a1656791/Th1_ZEB2_KO/070619/steve_script
31.
32. ## Directories for Initial FastQC
33. RAWDATA=${PROJROOT}/0_rawData
34. # mkdir -p ${RAWDATA}/FastQC
35.
36. ## Setup for Trimmed data
37. TRIMDATA=${PROJROOT}/1_trimmedData
38. mkdir -p ${TRIMDATA}/fastq
39. mkdir -p ${TRIMDATA}/FastQC
40. mkdir -p ${TRIMDATA}/logs
41.
42. ## Setup for genome alignment
43. ALIGNDATA=${PROJROOT}/2_alignedData
44. mkdir -p ${ALIGNDATA}/logs
45. mkdir -p ${ALIGNDATA}/bams
46. mkdir -p ${ALIGNDATA}/FastQC
47. mkdir -p ${ALIGNDATA}/featureCounts
48.
49. ##-----##
50. ## FastQC on the raw data (not required)
51. ##-----##
52.

```

```

53. # fastqc -t ${CORES} -o ${RAWDATA}/FastQC --noextract
    ${RAWDATA}/fastq/*fastq.gz
54.
55. ##-----##
56. ## Trimming the Merged data
57. ##-----##
58.
59. for R1 in ${RAWDATA}/fastq/*R1.fastq.gz
60.     do
61.
62.         echo -e "Currently working on ${R1}"
63.
64.         # Now create the output filenames
65.         out1=${TRIMDATA}/fastq/${basename $R1}
66.         BNAME=${TRIMDATA}/fastq/${basename ${R1%_R1.fastq.gz}}
67.         echo -e "Output file 1 will be ${out1}"
68.
69.         ## Repeat for R2 reads
70.         R2=${R1%_R1.fastq.gz}_R2.fastq.gz
71.         out2=${out1%_R1.fastq.gz}_R2.fastq.gz
72.
73.         echo -e "Trimming:\t${R1}\n\t${R2}"
74.         # Trim
75.         AdapterRemoval \
76.             --gzip \
77.             --trimns \
78.             --trimqualities \
79.             --minquality 30 \
80.             --minlength 50 \
81.             --threads ${CORES} \
82.             --basename ${BNAME} \
83.             --output1 ${out1} \
84.             --output2 ${out2} \
85.             --file1 ${R1} \
86.             --file2 ${R2}
87.
88.     done
89.
90. # Move the log files into their own folder
91. mv ${TRIMDATA}/fastq/*settings ${TRIMDATA}/logs
92.
93. # Run FastQC
94. fastqc -t ${CORES} -o ${TRIMDATA}/FastQC --noextract
    ${TRIMDATA}/fastq/*fastq.gz
95.
96. ##-----##
97. ## Aligning trimmed data to the genome
98. ##-----##
99.
100. ## Aligning, filtering and sorting
101. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
102.     do
103.
104.         R2=${R1%_R1.fastq.gz}_R2.fastq.gz
105.         BNAME=$(basename ${R1%_R1.fastq.gz})
106.         echo -e "STAR will align:\t${R1}\n\t${R2}"
107.
108.         STAR \
109.             --runThreadN ${CORES} \
110.             --genomeDir ${REFS}/star \
111.             --readFilesIn ${R1} ${R2} \

```



```

112.         --readFilesCommand gunzip -c \
113.         --outFileNamePrefix ${ALIGNDATA}/bams/${BNAME} \
114.         --outSAMtype BAM SortedByCoordinate
115.
116.     done
117.
118. # Move the log files into their own folder
119. mv ${ALIGNDATA}/bams/*out ${ALIGNDATA}/logs
120. mv ${ALIGNDATA}/bams/*tab ${ALIGNDATA}/logs
121.
122. # Fastqc and indexing
123. for BAM in ${ALIGNDATA}/bams/*.bam
124. do
125.     fastqc -t ${CORES} -f bam_mapped -o ${ALIGNDATA}/FastQC --
        noextract $(Bamias et al.)
126.     samtools index $(Bamias et al.)
127. done
128.
129. ##-----##
130. ## featureCounts
131. ##-----##
132.
133. ## Feature Counts - obtaining all sorted bam files
134. sampleList=`find ${ALIGNDATA}/bams -name "*out.bam" | tr '\n' ' '`
135.
136. ## Running featureCounts on the sorted bam files
137. featureCounts -Q 10 \
138.     -p \
139.     -s 2 \
140.     -T ${CORES} \
141.     -a ${GTF} \
142.     -o ${ALIGNDATA}/featureCounts/counts.out ${sampleList}
143.
144. ## Storing the output in a single file
145. cut -f1,7- ${ALIGNDATA}/featureCounts/counts.out | \
146. sed 1d > ${ALIGNDATA}/featureCounts/genes.out
147.
148.
149. ##-----##
150. ## kallisto
151. ##-----##
152.
153. ## Just submit these as independent jobs at this point as we don't
        need to do any
154. ## post alignment QC now
155. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
156. do
157.     sbatch ${PROJROOT}/bash/kallistoSingle.sh ${R1}
158. done
159.
160. ##-----##
161. ## salmon
162. ##-----##
163.
164. ## Just submit these as independent jobs at this point as we don't
        need to do any
165. ## post alignment QC now
166. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
167. do
168.     sbatch ${PROJROOT}/bash/salmonSingle.sh ${R1}
169. done

```

## Additional script for Salmon

```
1. #!/bin/bash
2. #SBATCH -p batch
3. #SBATCH -N 1
4. #SBATCH -n 1
5. #SBATCH --time=8:30:00
6. #SBATCH --mem=4GB
7. #SBATCH --mail-type=END
8. #SBATCH --mail-type=FAIL
9. #SBATCH --mail-user=soonwei.wong@adelaide.edu.au
10.
11. ## Cores
12. CORES=16
13.
14. # Load modules
15. module load Anaconda3/2019.03
16. source activate SalmonEnv
17. module load SAMtools/1.10-foss-2016b
18.
19. ## Genomic Data Files
20. IDX=/fast/users/a1656791/Salmon/salmon_index
21.
22. ## Directories
23. PROJROOT=/fast/users/a1656791/Th1_ZEB2_KO/070619/steve_script
24. TRIMDATA=${PROJROOT}/1_trimmedData
25.
26. ## Setup for salmon output
27. ALIGNDATA=${PROJROOT}/4_salmon
28.
29. ## Now organise the input files
30. F1=$1
31. F2=${F1%R1.fastq.gz}R2.fastq.gz
32.
33. ## Organise the output files
34. OUTDIR=${ALIGNDATA}/${basename ${F1%_R1.fastq.gz}}
35. echo -e "Creating ${OUTDIR}"
36. mkdir -p ${OUTDIR}
37. OUTBAM=${OUTDIR}/${basename ${F1%_R1.fastq.gz}.bam}
38.
39. echo -e "Currently aligning:\n\t${F1}\n\t${F2}"
40. echo -e "Output will be written to ${OUTDIR}"
41.
42. ## Check salmon version
43. salmon --version
44.
45. ## Transcript quantification with salmon quant
46. salmon quant \
47.     -i ${IDX} \
48.     -l A \
49.     -1 ${F1} -2 ${F2} \
50.     --useVBOpt \
51.     --seqBias \
52.     --numBootstraps 50 \
53.     --validateMappings \
54.     --threads ${CORES} \
55.     -o ${OUTDIR}
```

### 7.3.3 Bash script for activated Th1 ATAC-seq

```

1. #!/bin/bash
2. #SBATCH -p batch
3. #SBATCH -N 1
4. #SBATCH -n 16
5. #SBATCH --time=06:15:00
6. #SBATCH --mem=48GB
7. #SBATCH --mail-type=END
8. #SBATCH --mail-type=FAIL
9. #SBATCH --mail-user=soonwei.wong@adelaide.edu.au
10.
11. ## Note that the FastQC reports on the raw data were previously
    generated
12. ## for a preliminary meeting
13.
14. ## Cores
15. CORES=16
16.
17. ## Modules
18. module load FastQC/0.11.7
19. module load Anaconda3/2019.03
20. module load AdapterRemoval/2.2.1-foss-2016b
21. source activate ATACenv
22. module load SAMtools/1.10-foss-2016b
23. conda list
24.
25. ## Genomic Data Files
26. REFS=/data/biorefs/reference_genomes/ensembl-release-
    98/homo_sapiens
27. IDX=/fast/users/a1656791/ATAC_Th1/Index/Homo_sapiens/UCSC/hg38/Sequ
    ence/Bowtie2Index/genome
28. BLACKLIST=/fast/users/a1656791/ATAC_Th1/hg38-blacklist.v2.bed.gz
29.
30. ## Directories
31. PROJROOT=/fast/users/a1656791/ATAC_Th1
32.
33. ## Directories for Initial FastQC
34. RAWDATA=${PROJROOT}/0_rawData
35. # mkdir -p ${RAWDATA}/FastQC
36.
37. ## Setup for Trimmed data
38. TRIMDATA=${PROJROOT}/1_trimmedData
39. mkdir -p ${TRIMDATA}/fastq
40. mkdir -p ${TRIMDATA}/FastQC
41. mkdir -p ${TRIMDATA}/logs
42.
43. ## Setup for genome alignment
44. ALIGNEDATA=${PROJROOT}/2_alignedData
45. mkdir -p ${ALIGNEDATA}/logs
46. mkdir -p ${ALIGNEDATA}/bams
47. mkdir -p ${ALIGNEDATA}/FastQC
48. mkdir -p ${ALIGNEDATA}/Genrich
49.
50.
51. ##-----##
52. ## FastQC on the raw data (not required)
53. ##-----##
54.

```

```

55. # fastqc -t ${CORES} -o ${RAWDATA}/FastQC --noextract
    ${RAWDATA}/fastq/*fastq
56.
57.
58. ##-----##
59. ## Aligning trimmed data to the genome
60. ##-----##
61.
62. ## Aligning, filtering and sorting
63. for R1 in ${TRIMDATA}/fastq/*R1.fastq.gz
64.     do
65.
66.         R2=${R1%_R1.fastq.gz}_R2.fastq.gz
67.         BNAME=$(basename ${R1%_R1.fastq.gz})
68.         echo -e "Bowtie2 will align:\t${R1}\n\t${R2}"
69.
70.         bowtie2 \
71.             -p ${CORES} \
72.             --sensitive \
73.             -X 2000 \
74.             -k 20 \
75.             -x ${IDX} \
76.             -1 ${R1} \
77.             -2 ${R2} \
78.             | samtools view -u - \
79.             | samtools sort -n -o ${ALIGNDATA}/bams/${BNAME}.bam -
80.
81.     done
82.
83. # Move the log files into their own folder
84. #mv ${ALIGNDATA}/bams/*out ${ALIGNDATA}/logs
85. #mv ${ALIGNDATA}/bams/*tab ${ALIGNDATA}/logs
86.
87. # Fastqc and indexing
88. #for BAM in ${ALIGNDATA}/bams/*.bam
89. # do
90. #     fastqc -t ${CORES} -f bam_mapped -o ${ALIGNDATA}/FastQC --
    noextract $(Bamias et al.)
91. # done
92.
93. ##-----##
94. ## Peak Calling
95. ##-----##
96.
97. ## Obtaining all sorted bam files
98. sampleList=`find ${ALIGNDATA}/bams -name "*.bam" | tr '\n' ',' | sed
    's/./$//'\`
99. #echo -e "${sampleList} for peak calling"
100. ## Running Genrich on the sorted bam files
101.
102.     Genrich \
103.         -t ${sampleList} \
104.         -o ${ALIGNDATA}/Genrich/combined_peak.out \
105.         -f combined.log \
106.         -j \
107.         -r \
108.         -E ${BLACKLIST} \
109.         -e chrM,chrUn,random \
110.         -v \
111.         -b ${ALIGNDATA}/Genrich/combined.bed

```

## 7.4 RNA-seq R Scripts

### 7.4.1 Differential Expression Analysis

```
1. ---
2. title: "ZEB2 KO in Th1 EM: Differential Expression Analysis"
3. author: "Vincent Wong"
4. date: "`r format(Sys.time(), '%d %B, %Y')`"
5. output:
6.   html_document:
7.     fig_caption: yes
8.     fig_height: 6
9.     fig_width: 8
10.    toc: yes
11. editor_options:
12.   chunk_output_type: inline
13. ---
14.
15. ```{r setup, include=FALSE}
16. knitr::opts_chunk$set(
17.   autodep = TRUE,
18.   echo = TRUE,
19.   warning = FALSE,
20.   message = FALSE,
21.   fig.align = "center")
22. ```
23.
24. # Setup
25. First we load all the required `R` packages
26. ```{r loadPackages}
27. library(ngsReports)
28. library(tidyverse)
29. library(magrittr)
30. library(edgeR)
31. library(RUVSeq)
32. library(AnnotationHub)
33. library(ensemblDb)
34. library(scales)
35. library(cqn)
36. library(pander)
37. library(ggrepel)
38. library(pheatmap)
39. library(RColorBrewer)
40. library(goseq)
41. library(msigdbR)
42. library(AnnotationDbi)
43. ```
44.
45. ```{r setOpts}
46. panderOptions("table.split.table", Inf)
47. panderOptions("table.style", "markdown")
48. panderOptions("big.mark", ",")
49. ```
50.
51. ## Annotations
52.
53. ```{r annotationSetup}
```

```

54. ah <- AnnotationHub() %>%
55.   subset(species == "Homo sapiens") %>%
56.   subset(dataprovider == "Ensembl") %>%
57.   subset(genome == "GRCh38") %>%
58.   subset(rdaclass == "EnsDb")
59. ensDb <- ah[["AH79689"]]
60. ```
61.
62. ```{r transAnnotation}
63. grTrans <- transcripts(ensDb)
64. trLengths <- exonsBy(ensDb, "tx") %>%
65.   width() %>%
66.   vapply(sum, integer(1))
67. mcols(grTrans)$length <- trLengths[names(grTrans)]
68. ```
69.
70. ```{r geneAnnotation}
71. gcGene <- grTrans %>%
72.   mcols() %>%
73.   as.data.frame() %>%
74.   dplyr::select(gene_id, tx_id, gc_content, length, tx_name) %>%
75.   as_tibble() %>%
76.   group_by(gene_id) %>%
77.   summarise(
78.     gc_content = sum(gc_content*length) / sum(length),
79.     length = ceiling(median(length)),)
80.
81. grGenes <- genes(ensDb)
82. mcols(grGenes) %<>%
83.   as.data.frame() %>%
84.   left_join(gcGene) %>%
85.   as.data.frame() %>%
86.   DataFrame()
87. ```
88.
89. Similarly to the Quality Assessment steps, `GRanges` objects were
    formed at the gene and transcript levels, to enable estimation of GC
    content and length for each transcript and gene.
90. GC content and transcript length are available for each transcript,
    and for gene-level estimates, GC content was taken as the sum of all
    GC bases divided by the sum of all transcript lengths, effectively
    averaging across all transcripts.
91. Gene length was defined as the median transcript length.
92.
93. ```{r samplesAndLabels}
94. setwd("C:/Users/cronos/Box Sync/Desktop/070619/steve_script")
95. samples <- read_csv("samples.csv") %>%
96.   distinct(sampleName, .keep_all = TRUE) %>%
97.   dplyr::select(sample = sampleName, sampleID,
98.     group=genotype,label,memory,batch, replicate) %>%
99.   mutate(
100.     group = factor(group, levels = c("WT", "KO")),
101.     sampleID = factor(sampleID))
102. genoCols <- samples$group %>%
103.   levels() %>%
104.   length() %>%
105.   brewer.pal("Set1") %>%
106.   setNames(levels(samples$group)) %>%
107.   .[-3] #remove 3rd colour
108.

```

```

109. donorCols <- samples$sampleID %>%
110.   as.factor() %>%
111.   levels() %>%
112.   length() %>%
113.   brewer.pal("Set2") %>%
114.   setNames(levels(samples$sampleID))
115.   ```
116.
117. Sample metadata was also loaded, with only the sampleID and
   genotype being retained.
118. All other fields were considered irrelevant.
119.
120. ## Count Data
121. Load the transcript-level counts
122. ```{r transCounts}
123. # transCounts <- list.files("C:/Users/cronos/Box
   Sync/Desktop/070619/steve_script/3_kallisto_ensdb98", full.names =
   TRUE) %>%
124. transCounts <- list.files("C:/Users/cronos/Box
   Sync/Desktop/070619/4_salmon", full.names = TRUE) %>%
125.   catchSalmon()
126. colnames(transCounts$counts) %<>%
127.   basename() %>%
128.   str_remove("_combined")
129. libSizes <- colSums(transCounts$counts)
130.   ```
131.
132. ```{r dgeList Gene}
133. minCPM <- 1
134. minSamples <- 4
135. dgeList <-
   transCounts$counts[,grepl("EM",colnames(transCounts$counts))] %>%
   #select EM only
136. # .[,!grepl("18",colnames(.))] %>% #remove donor 18
137. as.data.frame() %>%
138. rownames_to_column("tx_id_version") %>%
139. gather(key = "sample", value = "Count", -tx_id_version) %>%
140. dplyr::filter(Count > 0) %>%
141. left_join(grTrans[,c("tx_id_version", "gene_id", "tx_id")] %>%
142.   mcols() %>%
143.   as.data.frame()) %>%
144. group_by(gene_id, sample) %>%
145. summarise(Count = sum(Count)) %>%
146. spread(sample, Count, fill = 0) %>%
147. as.data.frame() %>%
148. column_to_rownames("gene_id") %>%
149. .[rowSums(cpm(.) >= minCPM) >= minSamples,] %>%
150. DGEList(
151.   samples = tibble(sample = colnames(.)) %>%
152.     left_join(samples),
153.   genes = grGenes[rownames(.)] %>%
154.     as.data.frame() %>%
155.     dplyr::select(
156.       chromosome = seqnames, start, end,
157.       gene_id, gene_name, gene_biotype, description,
158.       entrezid, gc_content, length) %>%
159. #.[grepl("protein_coding", .$genes$gene_biotype),] %>%
160. .[!grepl("rRNA", .$genes$gene_biotype),] %>%
161. .[!grepl("lncRNA", .$genes$gene_biotype),] %>%
162.   calcNormFactors()
163.   ```

```

```

164.
165. Gene-level count data as output by `salmon`, was loaded and formed
    into a `DGEList` object.
166. During this process, genes were removed if:
167.
168. ```{r plotDensities, fig.width=5, fig.height=4,
    fig.cap="*Expression density plots for all samples after filtering,
    showing logCPM values.*"}
169.
170. dgeList %>%
171.   cpm(log = TRUE) %>%
172.   as.data.frame() %>%
173.   pivot_longer(
174.     cols = everything(),
175.     names_to = "sample",
176.     values_to = "logCPM"
177.   ) %>%
178.   split(f = .$sample) %>%
179.   lapply(function(x) {
180.     d <- density(x$logCPM)
181.     tibble(
182.       sample = unique(x$sample),
183.       x = d$x,
184.       y = d$y
185.     )
186.   }) %>%
187.   bind_rows() %>%
188.   left_join(samples) %>%
189.   ggplot(aes(x, y, colour = group, group = sample)) +
190.   geom_line() +
191.   scale_colour_manual(
192.     values = genoCols
193.   ) +
194.   labs(
195.     x = "logCPM",
196.     y = "Density",
197.     colour = "Group"
198.   )
199. ```
200.
201. Initial library sizes were between `r libSizes %>% range() %>%
    comma() %>% pander()`
202. This dataset retained `r comma(nrow(dgeList))` genes for DGE
    analysis.
203.
204. ```{r plotLibSize, echo=FALSE, fig.cap = "*Library Sizes after
    removal of undetectable genes and summarising to gene-level counts.
    Median library size is shown as the dashed line.*"}
205. dgeList$samples %>%
206.   ggplot(aes(sample, lib.size, fill = group)) +
207.   geom_bar(stat = "identity") +
208.   geom_hline(
209.     aes(yintercept = lib.size),
210.     data = . %>% summarise_at(vars(lib.size), median),
211.     linetype = 2
212.   ) +
213.   scale_y_continuous(expand = expand_scale(c(0, 0.05)), labels
    = comma) +
214.   scale_fill_manual(values = genoCols) +
215.   labs(x = "Sample", y = "Library Size", fill = "Group") +
216.   facet_wrap(~ group,

```



```

217.         scales = "free_x",
218.         labeller = as_labeller(c(WT = "Wildtype", KO =
      "Knockout")))) +
219.     theme(axis.text.x = element_text(angle = 90))
220. ```
221.
222. ### Additional Functions
223.
224. ```{r labellers}
225. # contLabeller <- as_labeller(
226. #   c(KOVsWT = "-/- Vs +/+)",
227. #     KO = "-/-",
228. #     WT = "+/+"))
229.
230. contLabeller <- as_labeller(
231.   c(KO = "KO Vs WT"))
232.
233. geneLabeller <- structure(grGenes$gene_name, names =
      grGenes$gene_id) %>%
234.   as_labeller()
235. ```
236.
237. Labeller functions for genotypes, contrasts and gene names were
      additionally defined for simpler plotting using `ggplot2`.
238.
239. # Analysis
240.
241. ## Initial PCA
242.
243. ```{r pca}
244. pca <- dgeList %>%
245.   cpm(log = TRUE) %>%
246.   t() %>%
247.   prcomp()
248. pcaVars <- percent_format(0.1)(summary(pca)$importance["Proportion
      of Variance",])
249. ```
250.
251. ### PCA1/PCA2
252.
253. ```{r plotPCA1/2, fig.cap="*PCA of gene-level counts.*"}
254. pca$x %>%
255.   as.data.frame() %>%
256.   rownames_to_column("sample") %>%
257.   as_tibble() %>%
258.   dplyr::select(sample, PC1, PC2) %>%
259.   left_join(dgeList$samples, "sample") %>%
260.   ggplot(aes(x=PC1, y=PC2, colour = group)) +
261.   geom_point(aes(shape=as.factor(sampleID)), alpha = 0.7, size = 3)
      +
262.   geom_text_repel(aes(label = label), size=5.4, show.legend =
      FALSE) +
263.   stat_ellipse(geom = "polygon", alpha = 0.1, show.legend = FALSE,
      aes(fill=group)) +
264.   #scale_shape_manual(values=c(15,11,19,17))+ #change shape
265.   scale_colour_manual(
266.     values = genoCols) +
267.   labs(
268.     x = paste0("PC1 (", pcaVars[["PC1"]], ")"),
269.     y = paste0("PC2 (", pcaVars[["PC2"]], ")"),
270.     colour = "Group",

```

```

271.     shape = "Donor")
272.   ```
273.
274. ### PCA3/PCA4
275. ```{r plotPCA3/4, fig.cap="*PCA of gene-level counts.*"}
276. pca$x %>%
277.   as.data.frame() %>%
278.   rownames_to_column("sample") %>%
279.   as_tibble() %>%
280.   dplyr::select(sample, PC3, PC4) %>%
281.   left_join(dgeList$samples,"sample") %>%
282.   ggplot(aes(x=PC3, y=PC4, colour = group)) +
283.   geom_point(aes(shape=as.factor(sampleID)),alpha = 0.7, size = 3)
284.   +
285.   geom_text_repel(aes(label = label), size=5.4, show.legend =
286.     FALSE) +
287.   stat_ellipse(geom = "polygon", alpha = 0.1, show.legend = FALSE,
288.     aes(fill=group)) +
289.   #scale_shape_manual(values=c(15,11,19,17))+ #change shape
290.   scale_colour_manual(
291.     values = genoCols) +
292.   labs(
293.     x = paste0("PC3 (", pcaVars[["PC3"]], ")"),
294.     y = paste0("PC4 (", pcaVars[["PC4"]], ")"),
295.     colour = "Group",
296.     shape = "Donor")
297.   ```
298.
299.
300. A Principal Component Analysis (PCA) was also performed using
301. logCPM values from each sample.
302. Both WT and KO appear to cluster together based on Donor.
303.
304.
305.
306. ## 1st Differential Expression
307. ```{r}
308. # Create model matrix
309. design <- model.matrix(~group, data = dgeList$samples)
310. ```
311.
312.
313.
314. # # Perform exact test on DGEList
315. # topTable <- dgeList %>%
316. #   estimateDisp(design = design) %>%
317. #   exactTest() %>%
318. #   topTags(n = Inf) %>%
319. #   .$table %>%
320. #   as_tibble() %>%
321. #   rownames_to_column("no.") %>%
322. #   mutate(DE = FDR < 0.05)
323. # #no DE
324. ```
325.
326.
327.
328.
329.
330.
331.
332.
333.
334.
335.
336.
337.
338.
339.
340.
341.
342.
343.
344.
345.
346.
347.
348.
349.
350.
351.
352.
353.
354.
355.
356.
357.
358.
359.
360.
361.
362.
363.
364.
365.
366.
367.
368.
369.
370.
371.
372.
373.
374.
375.
376.
377.
378.
379.
380.
381.
382.
383.
384.
385.
386.
387.
388.
389.
390.
391.
392.
393.
394.
395.
396.
397.
398.
399.
400.
401.
402.
403.
404.
405.
406.
407.
408.
409.
410.
411.
412.
413.
414.
415.
416.
417.
418.
419.
420.
421.
422.
423.
424.
425.
426.
427.
428.
429.
430.
431.
432.
433.
434.
435.
436.
437.
438.
439.
440.
441.
442.
443.
444.
445.
446.
447.
448.
449.
450.
451.
452.
453.
454.
455.
456.
457.
458.
459.
460.
461.
462.
463.
464.
465.
466.
467.
468.
469.
470.
471.
472.
473.
474.
475.
476.
477.
478.
479.
480.
481.
482.
483.
484.
485.
486.
487.
488.
489.
490.
491.
492.
493.
494.
495.
496.
497.
498.
499.
500.
501.
502.
503.
504.
505.
506.
507.
508.
509.
510.
511.
512.
513.
514.
515.
516.
517.
518.
519.
520.
521.
522.
523.
524.
525.
526.
527.
528.
529.
530.
531.
532.
533.
534.
535.
536.
537.
538.
539.
540.
541.
542.
543.
544.
545.
546.
547.
548.
549.
550.
551.
552.
553.
554.
555.
556.
557.
558.
559.
560.
561.
562.
563.
564.
565.
566.
567.
568.
569.
570.
571.
572.
573.
574.
575.
576.
577.
578.
579.
580.
581.
582.
583.
584.
585.
586.
587.
588.
589.
590.
591.
592.
593.
594.
595.
596.
597.
598.
599.
600.
601.
602.
603.
604.
605.
606.
607.
608.
609.
610.
611.
612.
613.
614.
615.
616.
617.
618.
619.
620.
621.
622.
623.
624.
625.
626.
627.
628.
629.
630.
631.
632.
633.
634.
635.
636.
637.
638.
639.
640.
641.
642.
643.
644.
645.
646.
647.
648.
649.
650.
651.
652.
653.
654.
655.
656.
657.
658.
659.
660.
661.
662.
663.
664.
665.
666.
667.
668.
669.
670.
671.
672.
673.
674.
675.
676.
677.
678.
679.
680.
681.
682.
683.
684.
685.
686.
687.
688.
689.
690.
691.
692.
693.
694.
695.
696.
697.
698.
699.
700.
701.
702.
703.
704.
705.
706.
707.
708.
709.
710.
711.
712.
713.
714.
715.
716.
717.
718.
719.
720.
721.
722.
723.
724.
725.
726.
727.
728.
729.
730.
731.
732.
733.
734.
735.
736.
737.
738.
739.
740.
741.
742.
743.
744.
745.
746.
747.
748.
749.
750.
751.
752.
753.
754.
755.
756.
757.
758.
759.
760.
761.
762.
763.
764.
765.
766.
767.
768.
769.
770.
771.
772.
773.
774.
775.
776.
777.
778.
779.
780.
781.
782.
783.
784.
785.
786.
787.
788.
789.
790.
791.
792.
793.
794.
795.
796.
797.
798.
799.
800.
801.
802.
803.
804.
805.
806.
807.
808.
809.
810.
811.
812.
813.
814.
815.
816.
817.
818.
819.
820.
821.
822.
823.
824.
825.
826.
827.
828.
829.
830.
831.
832.
833.
834.
835.
836.
837.
838.
839.
840.
841.
842.
843.
844.
845.
846.
847.
848.
849.
850.
851.
852.
853.
854.
855.
856.
857.
858.
859.
860.
861.
862.
863.
864.
865.
866.
867.
868.
869.
870.
871.
872.
873.
874.
875.
876.
877.
878.
879.
880.
881.
882.
883.
884.
885.
886.
887.
888.
889.
890.
891.
892.
893.
894.
895.
896.
897.
898.
899.
900.
901.
902.
903.
904.
905.
906.
907.
908.
909.
910.
911.
912.
913.
914.
915.
916.
917.
918.
919.
920.
921.
922.
923.
924.
925.
926.
927.
928.
929.
930.
931.
932.
933.
934.
935.
936.
937.
938.
939.
940.
941.
942.
943.
944.
945.
946.
947.
948.
949.
950.
951.
952.
953.
954.
955.
956.
957.
958.
959.
960.
961.
962.
963.
964.
965.
966.
967.
968.
969.
970.
971.
972.
973.
974.
975.
976.
977.
978.
979.
980.
981.
982.
983.
984.
985.
986.
987.
988.
989.
990.
991.
992.
993.
994.
995.
996.
997.
998.
999.
1000.

```

```

328.   as_tibble() %>%
329.   dplyr::select(
330.     gene_id, gene_name, logFC, logCPM, PValue, FDR,
331.     everything()
332.   ) %>%
333.   rownames_to_column("no.") %>%
334.   mutate(DE = FDR < 0.05)
335.   #no DE
336.   ```
337.   ## RUVg Normalisation
338.
339.   If no genes are known a priori not to be influenced by the
340.   covariates of interest, one can obtain a set of "in-silico empirical"
341.   negative controls, e.g., least significantly DE genes based on a
342.   first-pass DE analysis performed prior to RUVg normalization.
343.   ```{r}
344.   empirical <- topTable %>%
345.     dplyr::arrange(desc(PValue)) %>%
346.     .[1:5000,] %>%
347.     .$gene_id
348.   ```
349.   Apply RUVg using empirical control genes, k=1 is better
350.   ```{r RUVg}
351.   # Run RUVSeq
352.   RUVg <- RUVg(dgeList$counts, empirical, 3)
353.
354.   dgeList$normalizedCounts <- RUVg$normalizedCounts
355.   ```
356.   ## PCA
357.   ### RUVg normalised PCA1/PCA2
358.   ```{r PCA, fig.cap = "*Principal component analysis showed
359.   separation across PC1 between condition accounting for 23% of the
360.   variation, indicated by colour of the points.*", fig.height=4,
361.   fig.width=6}
362.   pca <- RUVg$normalizedCounts %>%
363.     cpm(log = TRUE) %>%
364.     t() %>%
365.     prcomp()
366.   pcaVars <- percent_format(0.1)(summary(pca)$importance["Proportion
367.   of Variance",])
368.
369.   pca$x %>%
370.     as.data.frame() %>%
371.     rownames_to_column("sample") %>%
372.     as_tibble() %>%
373.     dplyr::select(sample, PC1, PC2) %>%
374.     left_join(dgeList$samples,"sample") %>%
375.     ggplot(aes(x=PC1, y=PC2, colour = group)) +
376.     geom_point(aes(shape=as.factor(sampleID)),alpha = 0.7, size = 3)
377.     +
378.     geom_text_repel(aes(label = label), size=5.4, show.legend =
379.     FALSE) +
380.     stat_ellipse(geom = "polygon", alpha = 0.1, show.legend = FALSE,
381.     aes(fill=group)) +
382.     #scale_shape_manual(values=c(15,11,19,17))+ #change shape
383.     scale_colour_manual(values = genoCols) +

```

```

378. labs(
379.   x = paste0("PC1 (", pcaVars[["PC1"]], ")"),
380.   y = paste0("PC2 (", pcaVars[["PC2"]], ")"),
381.   colour = "Genotype",
382.   shape = "Donor")
383. ```
384.
385. ## DE Analysis
386.
387. ```{r}
388. # Create copy of DGE List
389. dgeRUVg <- dgeList
390. # Bind W_1 from RUVSeq analysis to $samples
391. dgeRUVg$samples %<>% cbind(RUVg$W)
392. # Create design matrix
393. design <- model.matrix(~group + W_1 + W_2 + W_3 , data =
  dgeRUVg$samples)
394.
395. dgeRUVg %<>% estimateGLMCommonDisp(design) %>%
396.   estimateGLMTagwiseDisp(design)
397. ```
398.
399. ```{r foldChange & significance}
400. minLfc <- log2(1.1)
401. alpha <- 0.05
402. ```
403.
404. ```{r}
405. # Perform DE analysis
406. topTableRUVg <- glmFit(dgeRUVg, design) %>%
407.   glmTreat (coef=2, lfc=minLfc) %>%
408.   topTags(n = Inf) %>%
409.   .$table %>%
410.   as_tibble() %>%
411.   dplyr::select(chromosome,start, end,
412.     gene_id, gene_name, logFC, logCPM, PValue, FDR,
413.     everything()
414.   ) %>%
415.   mutate(DE = FDR < 0.05)
416.
417. topTableRUVg %<>%
418.   mutate(Status = case_when(
419.     DE ==TRUE & logFC < 0 ~ "Down",
420.     DE == TRUE & logFC > 0 ~ "Up"),
421.     Status = ifelse(is.na(Status), "NotSig", Status))
422.
423. a<-topTableRUVg
424. ```
425.
426. Using these criteria, the following initial DE gene-sets were
  defined:
427.
428. ```{r printInitialDE, results='asis'}
429. table(topTableRUVg$Status) %>%
430.   as.data.frame() %>%
431.   set_names(NULL)
432. ```
433.
434. ```{r}
435. topTableRUVgLRT <- glmFit(dgeRUVg, design) %>%

```

```

436. glmLRT (coef=2) %>%
437. topTags(n = Inf) %>%
438. .$table %>%
439. as_tibble() %>%
440. dplyr::select(
441.   gene_id, gene_name, logFC, logCPM, PValue, FDR,
     everything()
442. ) %>%
443.   rownames_to_column("no.") %>%
444.   mutate(DE = FDR < 0.05)
445.
446. topTableRUVgLRT %<>%
447.   mutate(Status = case_when(
448.     DE ==TRUE & logFC < 0 ~ "Down",
449.     DE == TRUE & logFC > 0 ~ "Up"),
450.     Status = ifelse(is.na(Status), "NotSig", Status))
451.
452. table(topTableRUVgLRT$Status) %>%
453.   as.data.frame() %>%
454.   set_names(NULL)
455. ```
456.
457.
458.
459. ```{r}
460. # Summary of DE genes
461. topTableRUVg %>%
462.   .[!grepl("HLA", .$gene_name),] %>%
463.   dplyr::slice(1:10) %>%
464.   dplyr::filter(FDR < 0.05) %>%
465.   dplyr::select(gene_id, gene_name, logCPM, logFC, PValue, FDR, DE,
     Status) %>%
466.   pander(style = "rmarkdown", split.tables = Inf)
467. ```
468. ```{r}
469. deCols <- c(
470.   'Up' = rgb(1, 0, 0, 1),
471.   'Down' = rgb(0, 0, 1, 1),
472.   'NotSig' = rgb(0.5, 0.5, 0.5, 0.7),
473.   `FALSE` = rgb(0.5, 0.5, 0.5, 0.7),
474.   `TRUE` = rgb(1, 0, 0, 0.7))
475. ```
476.
477.
478. ```{r}
479. topTableRUVg%>%
480.   bind_rows() %>%
481.   ggplot(aes(logFC, -log10(PValue), colour = Status)) +
482.   geom_point(alpha = 0.4) +
483.   geom_label_repel(
484.     aes(label = gene_name), size = 4,
485.     max.overlaps = getOption("ggrepel.max.overlaps", default = 30),
486.     data = . %>% .[c(1:30),])+
487.   geom_vline(
488.     xintercept = c(-1, 1)*minLfc,
489.     linetype = 2,
490.     colour = "red"
491.   ) +
492.   scale_colour_manual(values = deCols) +
493.   scale_x_continuous(breaks = seq(-4, 4, by = 2)) +
494.   theme(legend.position = "none")

```

```

495. ````
496.
497.
498. ````{r plotTop20, echo=FALSE, fig.cap="*Top 49 most DE genes, when
      sorting by p-value.*"}
499. topTableRUVg %>%
500.   dplyr::slice(1:20) %>%
501.   dplyr::select(gene_id) %>%
502.   cbind(cpm(dgeList$normalizedCounts)[.$gene_id,]) %>%
503.   as_tibble() %>%
504.   gather(key = "sample", value = "CPM", -gene_id) %>%
505.   mutate(CPM = CPM + 1) %>%
506.   left_join(dgeList$samples) %>%
507.   ggplot(aes(group, CPM, fill = group)) +
508.   stat_boxplot(geom='errorbar', linetype=1, width=0.5)+
509.   geom_boxplot() +
510.   scale_fill_manual(values=genoCols)+
511.   facet_wrap(
512.     ~gene_id,
513.     scales = "free_y",
514.     labeller = as_labeller(
515.       structure(dgeList$genes$gene_name, names =
516.         dgeList$genes$gene_id)
517.     )) +
518.   scale_y_log10() +
519.   labs(x = "Treatment Group",
520.        y = "CPM + 1") +
521.   theme(legend.position = "none")
522.
523. ## Genes Tracking with *ZEB2^KO^*
524.
525. ````{r deGenes}
526. deGenes <- topTableRUVg$gene_id
527. ````
528.
529. ````{r commonSummaryHeatmap, fig.height=15, fig.width=8 , fig.cap=
      "*The 100 most highly-ranked genes by FDR which are commonly
      considered DE between KO and WT samples. Plotted values are z-score
      based on normalised counts.*"}
530. n <- 100
531.
532. cpm(dgeRUVg$normalizedCounts, log=TRUE) %>%
533.   magrittr::extract(deGenes[1:n],) %>%
534.   set_rownames(unlist(geneLabeller(rownames(.)))) %>%
535.   t() %>%
536.   scale() %>%
537.   t() %>%
538.   pheatmap::pheatmap(
539.     color = viridis_pal(option = "magma")(100),
540.     border_color = NA,
541.     labels_col = dgeList$samples$label,
542.     legend_breaks = c(seq(-2, 2, by = 1)),
543.     legend_labels = c(seq(-2, 2, by = 1)),
544.     annotation_col = dgeList$samples %>%
545.       dplyr::select(Group = group, Donor = sampleID),
546.     annotation_names_col = FALSE,
547.     annotation_colors = list(Group = genoCols, Donor = donorCols))
548. ````
549.

```



## 7.4.2 Enrichment Analysis

### Gene Ontology and KEGG pathway Analysis

```

1. ---
2. title: "ZEB2 KO in Th1 EM: Enrichment Analysis"
3. author: "Vincent Wong"
4. date: "`r format(Sys.time(), '%d %B, %Y')`"
5. output: workflowr::wflow_html
6. editor_options:
7.   chunk_output_type: inline
8. ---
9.
10. ```{r setup, include=FALSE}
11. knitr::opts_chunk$set(
12.   autodep = TRUE,
13.   echo = TRUE,
14.   warning = FALSE,
15.   message = FALSE,
16.   fig.align = "center")
17. ```
18.
19. # Setup
20. First we load all the required `R` packages
21. ```{r loadPackages}
22. library(ngsReports)
23. library(tidyverse)
24. library(magrittr)
25. library(edgeR)
26. library(RUVSeq)
27. library(AnnotationHub)
28. library(ensemldb)
29. library(scales)
30. library(cqn)
31. library(pander)
32. library(ggrepel)
33. library(pheatmap)
34. library(RColorBrewer)
35. library(goseq)
36. library(msigdb)
37. library(AnnotationDbi)
38. ```
39.
40. ```{r setOpts}
41. panderOptions("table.split.table", Inf)
42. panderOptions("table.style", "rmarkdown")
43. panderOptions("big.mark", ",")
44. ```
45.
46. # Steve GO Enrichment Analysis
47. ## GO Enrichment Analysis with topTableLRT
48. ```{r}
49. library(org.Hs.eg.db)
50. library(enrichplot)
51. library(clusterProfiler)
52. library(ggplot2)
53. ```
54.
55. ```{r minPath}

```



```

56. minPath <- 4
57. goSummaries <-
  url("https://uofabioinformaticshub.github.io/summaries2GO/data/goSummaries.RDS") %>%
58.   readRDS() %>%
59.   dplyr::rename(go_id = id)
60.   ```
61.
62. A very brief enrichment analysis was then performed using the
  default settings of the function `goana()`.
63. As `goana` relies on entrezgene identifiers, analysis was switched
  to these IDs for this step of the analysis.
64. Results were then limited to GO terms with at least `r minPath`
  steps back the root terms for each ontology.
65.
66. # Gene list FDR < 0.05
67.
68. ## Gene Ontology
69. ```{r egIDS}
70. tt<-readRDS("C:/Users/cronos/Box
  Sync/Desktop/070619/topTableRUVg_log2_1.1_salmon.rds")
71. sigIDs0.05 <- tt %>%
72.   dplyr::filter(FDR < 0.05) %>%
73.   extract2("entrezid") %>%
74.   unlist() %>%
75.   na.omit() %>%
76.   unique()
77.
78. uv <- unlist(tt$entrezid) %>%
79.   na.omit() %>%
80.   unique()
81.   ```
82.
83. ```{r EM goRes, echo=TRUE, warning=TRUE}
84. goRes <- goana.default(sigIDs0.05, universe = uv, species =
  "Hs") %>%
85.   rownames_to_column("go_id") %>%
86.   dplyr::rename(ontology = Ont) %>%
87.   as_tibble() %>%
88.   dplyr::filter(DE > 0) %>%
89.   left_join(goSummaries)
90.   ```
91.
92. A total of "r comma(nrow(goRes))" GO terms were identified amongst
  the differentially expressed genes.
93.
94. ```{r topGORes}
95. goRes %>%
96.   dplyr::filter(shortest_path > minPath) %>%
97.   mutate(adjP = p.adjust(P.DE, "bonferroni"),
98.          FDR = p.adjust(P.DE, "fdr")) %>%
99.   arrange(ontology, P.DE) %>%
100.  dplyr::filter(FDR < 0.1) %>%
101.  mutate(Expected_DE = round(length(sigIDs0.05) * N /
  length(uv), 1)) %>%
102.  dplyr::select(`GO ID` = go_id, Term, Ontology = ontology, N,
  Obs_DE = DE, Expected_DE, adjP, FDR) %>%
103.  pander(caption = "*Significantly enriched GO terms using a
  FDR-adjusted p-value < 0.1 as the criteria for significance*",
104.         split.tables = Inf)
105.   ```

```

```

106. ```{r GO_barplot}
107. Enriched_GO <-goRes %>%
108.     dplyr::filter(shortest_path > minPath) %>%
109.     mutate(adjP = p.adjust(P.DE, "fdr"),
110.            FDR = p.adjust(P.DE, "fdr")) %>%
111.     arrange(ontology, P.DE) %>%
112.     dplyr::filter(adjP < 0.1) %>% mutate(
113.       GOTerm = paste(go_id, Term, sep = ":")
114.     )
115. # make a an ordered factor
116. Enriched_GO$Term <- factor(Enriched_GO$Term, levels =
117.   Enriched_GO$Term[order(Enriched_GO$ontology)])
118. Enriched_GO %>%
119.   ggplot(aes(x=DE,y=Term, fill = ontology)) +
120.   geom_bar(stat = "identity") +
121.   scale_y_discrete(label = function(x) stringr::str_trunc(x, 50)) +
122.   labs(x = "Number of Genes", y = "GO:Term") +
123.   scale_fill_discrete(name = "Ontology")
124.
125. ### GO Visualisations
126. ```{r}
127. go2Drop <- dplyr::filter(goSummaries, shortest_path <=
128.   minPath)$go_id
129. In order to visualise the results of the GO analysis, the R package
130. clusterProfiler was used. The same strategy was used for removal of
131. high-level GO terms, which resulted in the omission of r
132. comma(length(go2Drop)) high-level (i.e. less informative) GO terms
133. from the analysis.
134. ```{r}
135. orgDb <- org.Hs.eg.db
136. eGO <- c("BP", "CC", "MF") %>%
137.   sapply(function(x){
138.     enrichGO(
139.       gene = as.character(sigIDs0.05),
140.       universe = as.character(uv),
141.       ont = x,
142.       pAdjustMethod = "none", #plotting purposes not for
143.       stats
144.       pvalueCutoff = 0.1 ,
145.       minGSSize = 5,
146.       OrgDb = orgDb,
147.       readable = TRUE
148.     ) %>%
149.     dropGO(term = go2Drop)
150.   }, simplify = FALSE)
151.
152.
153.
154. ```{r fc}
155. fc <- structure(
156.   dplyr::filter(tt, DE)$logFC,
157.   names = dplyr::filter(tt, DE)$entrezid)
158.
159.
160.
161. ```{r fig.height=12, fig.width=12}
162. eGO$BP %>%
163.   #dropGO(term = c("GO:0071346")) %>% #to drop GO Term of interest
164.   cnetplot(foldChange = fc, showCategory = 10) +
165.   scale_colour_gradient2(low = "blue", mid = "white", high =
166.     "red") +

```

```

159.         guides(size = FALSE) +
160.         labs(colour = "logFC")
161.     ...
162.
163.     ```{r fig.height=6, fig.width=12}
164.     eGO$MF %>%
165.       #dropGO(term = c("GO:0071346")) %>% #to drop GO Term of interest
166.       cnetplot(foldChange = fc, showCategory = 1) +
167.       scale_colour_gradient2(low = "blue", mid = "white", high =
         "red") +
168.       guides(size = FALSE) +
169.       labs(colour = "logFC")
170.     ...
171.
172. ## KEGG Analysis
173. ```{r keggRes}
174. keggRes <- kegg.default(sigIDs0.05, universe = uv, species =
         "Hs") %>%
175.   rownames_to_column("kegg_id") %>%
176.   as_tibble() %>%
177.   dplyr::filter(DE > 0) %>%
178.   mutate(FDR = p.adjust(P.DE, method = "fdr")) %>%
179.   arrange(P.DE)
180.     ...
181.
182.     ```{r panderKeggRes}
183.     keggRes %>%
184.       #dplyr::filter(adjP < 0.05) %>%
185.       dplyr::filter(P.DE < 0.05) %>%
186.       mutate(Expected_DE = round(length(sigIDs0.05) * N /
         length(uv), 1)) %>%
187.       dplyr::select(`KEGG ID` = kegg_id, Pathway, N, Obs_DE = DE,
         Expected_DE, FDR, P.DE) %>%
188.       pander(caption = "*Significantly enriched KEGG terms using a
         Bonferroni-adjusted p-value < 0.1 as the criteria for significance,
         and using the more inclusive list of DE genes*",
         split.tables = Inf)
189.     ...
190.     ...
191.
192.     ```{r fig.height=10, fig.width=13}
193.     kk <- enrichKEGG(
194.       gene = as.character(sigIDs0.05),
195.       universe = as.character(uv),
196.       keyType="kegg",
197.       organism = 'hsa',
198.       pAdjustMethod = "none", #stats using goana for plotting
         purposes
199.       pvalueCutoff = 1,
200.       qvalueCutoff = 1)
201.     #convert geneid to SYMBOL
202.     kk <- setReadable(kk, orgDb, 'ENTREZID')
203.     #visualization
204.     cnetplot(kk, foldChange = fc, showCategory = 17) +
205.       scale_colour_gradient2(low = "blue", mid = "white", high =
         "red") +
206.       guides(size = FALSE) +
207.       labs(colour = "logFC")
208.
209.     # write.table(kk@result,file="kk.txt",
210.     #             row.names = TRUE, col.names = TRUE)
211.     ...

```

```

212.
213. # Gene list FDR < 0.1
214. ## Gene Ontology
215. ```{r egIDS}
216. tt<-readRDS("C:/Users/cronos/Box
  Sync/Desktop/070619/topTableRUVg_log2_1.1_salmon.rds")
217. tt %<>% mutate(DE2 = FDR < 0.1)
218.
219. sigIDs0.1 <- tt %>%
220.   dplyr::filter(FDR < 0.1) %>%
221.   extract2("entrezid") %>%
222.   unlist() %>%
223.   na.omit() %>%
224.   unique()
225.
226.
227. uv <- unlist(tt$entrezid) %>%
228.   na.omit() %>%
229.   unique()
230. ```
231.
232. ```{r EM goRes, echo=TRUE, warning=TRUE}
233. goRes <- goana.default(sigIDs0.1, universe = uv, species =
  "Hs") %>%
234.   rownames_to_column("go_id") %>%
235.   dplyr::rename(ontology = Ont) %>%
236.   as_tibble() %>%
237.   dplyr::filter(DE > 0) %>%
238.   left_join(goSummaries)
239. ```
240. ```
241.
242. A total of "r comma(nrow(goRes))" GO terms were identified amongst
  the differentially expressed genes.
243.
244. ```{r topGORes}
245. goRes %>%
246.   dplyr::filter(shortest_path > minPath) %>%
247.   mutate(FDR = p.adjust(P.DE, "fdr")) %>%
248.   arrange(ontology, P.DE) %>%
249.   dplyr::filter(FDR < 0.1) %>%
250.   mutate(Expected_DE = round(length(sigIDs0.1) * N /
  length(uv), 1)) %>%
251.   dplyr::select(`GO ID` = go_id, Term, Ontology = ontology, N,
  Obs_DE = DE, Expected_DE, P.DE, FDR) %>%
252.   pander(caption = "*Significantly enriched GO terms using a
  FDR-adjusted p-value < 0.1 as the criteria for significance*",
  split.tables = Inf)
253. ```
254. ```
255. ```{r GO_barplot}
256. Enriched_GO <-goRes %>%
257.   dplyr::filter(shortest_path > minPath) %>%
258.   mutate(adjP = p.adjust(P.DE, "fdr"),
  FDR = p.adjust(P.DE, "fdr")) %>%
259.   arrange(ontology, P.DE) %>%
260.   dplyr::filter(adjP < 0.1) %>% mutate(
  GOTerm = paste(go_id, Term, sep = ":"))
261.
262.
263.
264. # make a an ordered factor
265. Enriched_GO$Term <- factor(Enriched_GO$Term, levels =
  Enriched_GO$Term[order(Enriched_GO$ontology)])

```

```

266. Enriched_GO %>%
267.     ggplot(aes(x=DE,y=Term, fill = ontology)) +
268.     geom_bar(stat = "identity") +
269.     scale_y_discrete(label = function(x) stringr::str_trunc(x, 50)) +
270.     labs(x = "Number of Genes", y = "GO:Term") +
271.     scale_fill_discrete(name = "Ontology")
272. ```
273.
274. ### GO Visualisations
275. ```{r}
276. go2Drop <- dplyr::filter(goSummaries, shortest_path <=
  minPath)$go_id
277. ```
278. In order to visualise the results of the GO analysis, the R package
  clusterProfiler was used. The same strategy was used for removal of
  high-level GO terms, which resulted in the omission of r
  comma(length(go2Drop)) high-level (i.e. less informative) GO terms
  from the analysis.
279. ```{r}
280. orgDb <- org.Hs.eg.db
281. eGO <- c("BP", "CC", "MF") %>%
282.     sapply(function(x){
283.         enrichGO(
284.             gene = as.character(sigIDs0.1),
285.             universe = as.character(uv),
286.             ont = x,
287.             pAdjustMethod = "none",
288.             pvalueCutoff = 0.1,
289.             minGSSize = 5,
290.             OrgDb = orgDb,
291.             readable = TRUE
292.         ) %>%
293.             dropGO(term = go2Drop)
294.     }, simplify = FALSE)
295. ```
296. ```
297.
298. ```{r fc}
299. fc2 <- structure(
300.     dplyr::filter(tt, DE2)$logFC,
301.     names = dplyr::filter(tt, DE2)$entrezid)
302. ```
303.
304. ```{r fig.height=12, fig.width=10}
305. eGO$BP %>%
306.     cnetplot(foldChange = fc2, showCategory = 10) +
307.     scale_colour_gradient2(low = "blue", mid = "white", high =
  "red") +
308.     guides(size = FALSE) +
309.     labs(colour = "logFC")
310. ```
311.
312. ```{r}
313. eGO$MF %>%
314.     cnetplot(foldChange = fc2, showCategory = 7) +
315.     scale_colour_gradient2(low = "blue", mid = "white", high =
  "red") +
316.     guides(size = FALSE) +
317.     labs(colour = "logFC")
318. ```
319. ```{r}

```

```

320. eGO$CC %>%
321.     cnetplot(foldChange = fc2, showCategory = 2) +
322.     scale_colour_gradient2(low = "blue", mid = "white", high =
323.     "red") +
324.     guides(size = FALSE) +
325.     labs(colour = "logFC")
326.
327. ## KEGG Analysis
328. ```{r keggRes}
329. keggRes <- kegg.default(sigIDs0.1, universe = uv, species =
330.     "Hs") %>%
331.     rownames_to_column("kegg_id") %>%
332.     as_tibble() %>%
333.     dplyr::filter(DE > 0) %>%
334.     mutate(adjP = p.adjust(P.DE, method = "fdr")) %>%
335.     arrange(P.DE)
336.
337. ```{r panderKeggRes}
338. keggRes %>%
339.     #dplyr::filter(adjP < 0.05) %>%
340.     dplyr::filter(P.DE < 0.05) %>%
341.     mutate(Expected_DE = round(length(sigIDs0.1) * N /
342.     length(uv), 1)) %>%
343.     dplyr::select(`KEGG ID` = kegg_id, Pathway, N, Obs_DE = DE,
344.     Expected_DE, adjP, P.DE) %>%
345.     pander(caption = "*Significantly enriched KEGG terms using a
346.     Bonferroni-adjusted p-value < 0.1 as the criteria for significance,
347.     and using the more inclusive list of DE genes*",
348.     split.tables = Inf)
349.
350. ```{r fig.height=10, fig.width=13}
351. kk <- enrichKEGG(
352.     gene = as.character(sigIDs0.1),
353.     universe = as.character(uv),
354.     keyType="kegg",
355.     organism = 'hsa',
356.     pAdjustMethod = "fdr",
357.     pvalueCutoff = 1,
358.     qvalueCutoff = 1)
359. #convert geneid to SYMBOL
360. kk <- setReadable(kk, orgDb, 'ENTREZID')
361. #visualization
362. cnetplot(kk, foldChange = fc, showCategory = 20) +
363.     scale_colour_gradient2(low = "blue", mid = "white", high =
364.     "red") +
365.     guides(size = FALSE) +
366.     labs(colour = "logFC")
367.
368. ```{r}
369. final_productKEGG_Overrep <- heatmap(kk, foldChange = fc,
370.     showCategory = 14)+
371.     ggplot2::coord_flip() +ggplot2::ggtitle('KEGG Over-representation
372.     of Pathway Analysis ')+ggplot2::ylab('Annotations of KEGG Pathway')+
373.     ggplot2::xlab('Gene Symbols')
374.
375. final_productKEGG_Overrep
376.
377. ```

```

Gene set enrichment analysis

```

1. ---
2. title: "ZEB2 KO in Th1 EM: Gene Set Enrichment Analysis"
3. author: "Vincent Wong"
4. date: "`r format(Sys.time(), '%d %B, %Y')`"
5. output: workflowr::wflow_html
6. editor_options:
7.   chunk_output_type: inline
8. ---
9.
10. ```{r setup, include=FALSE}
11. knitr::opts_chunk$set(
12.   autodep = TRUE,
13.   echo = TRUE,
14.   warning = FALSE,
15.   message = FALSE,
16.   fig.align = "center")
17. ```
18.
19.
20. ```{r loadPackages}
21. library(tidyverse)
22. library(magrittr)
23. library(edgeR)
24. library(scales)
25. library(pander)
26. library(goseq)
27. library(msigdb)
28. library(RColorBrewer)
29. library(ngsReports)
30. library(UpSetR)
31. library(pheatmap)
32. ```
33.
34. ```{r}
35. dgeRUVg <- readRDS("C:/Users/cronos/Box
   Sync/Desktop/070619/dgeRUVg_Salmon.rds")
36. entrezGenes<-dgeRUVg$genes %>%
37.   dplyr::filter(!is.na(entrezid)) %>%
38.   unnest(entrezid) %>%
39.   dplyr::rename(entrez_gene = entrezid)
40.
41. topTableRUVg <- readRDS("C:/Users/cronos/Box
   Sync/Desktop/070619/topTableRUVg_log2_1.1_salmon.rds")
42.
43. deTable <- topTableRUVg %>%
44.   mutate(
45.     entrezid = dgeRUVg$genes$entrezid[gene_id])
46.
47. geneLabeller <- structure(topTableRUVg$gene_name, names =
   topTableRUVg$gene_id) %>%
48.   as_labeller()
49.
50. # Create design matrix
51. design <- model.matrix(~group, data = dgeRUVg$samples)
52.
53. cpm<-cpm(dgeRUVg$normalizedCounts,log = TRUE)
54. ```

```

```

55.
56. Gene Sets for Human:
57. H hallmark gene sets
58. C1 positional gene sets
59. C2 curated gene sets
60. C3 motif gene sets
61. C4 computational gene sets
62. C5 GO gene sets
63. C6 oncogenic signatures
64. C7 immunologic signatures
65.
66. ```{r}
67. minLfc=log2(1.1)
68. ```
69.
70. ```{r}
71. topTableRUVg <- readRDS("C:/Users/cronos/Box
  Sync/Desktop/070619/topTableRUVg_log2_1.1_salmon.rds")
72. topTableRUVg %>%
73.   ggplot(aes(logFC, -log10(PValue))) +
74.     geom_vline(xintercept = c(-1, 1), linetype = 2) +
75.     geom_point(aes(colour = DE), alpha = 0.6) +
76.     scale_colour_manual(values = c("grey20", "red"))
77.   ```
78.   ```{r}
79. topTableRUVg %>%
80.   ggplot(aes(logCPM, logFC)) +
81.     geom_hline(yintercept = c(-1, 1), linetype = 2) +
82.     geom_point(aes(colour = DE), alpha = 0.6) +
83.     geom_smooth(se = FALSE) +
84.     scale_colour_manual(values = c("grey20", "red"))
85.   ```
86.
87.   ```{r}
88. topTableRUVg %>%
89.   mutate(
90.     prot_coding = str_detect(gene_biotype, "protein_coding")
91.   ) %>%
92.   group_by(prot_coding, DE) %>%
93.   tally() %>%
94.   pivot_wider(
95.     id_cols = prot_coding, names_from = DE, names_prefix = "DE_",
96.     values_from = n
97.   )
98.   ```
99.   ```{r}
100. topTableRUVg %>%
101.   mutate(
102.     prot_coding = str_detect(gene_biotype, "protein_coding")
103.   ) %>%
104.   group_by(prot_coding, DE) %>%
105.   tally() %>%
106.   pivot_wider(
107.     id_cols = prot_coding, names_from = DE, names_prefix = "DE_",
108.     values_from = n
109.   ) %>%
110.   as.data.frame() %>%
111.   column_to_rownames("prot_coding") %>%
112.   fisher.test()

```



```

113. ```{r}
114. dgeRUVg <- readRDS("C:/Users/cronos/Box
  Sync/Desktop/070619/dgeRUVg_Salmon.rds")
115. ens98_GC <- dgeRUVg$genes %>% as.data.frame() %>%
  dplyr::select(gene_gc="gc_content", gene_length="length") %>%
  rownames_to_column("gene_id")
116. topTableRUVg <- left_join(topTableRUVg, ens98_GC)
117. ```
118.
119. ```{r}
120. deVec <- topTableRUVg$DE %>%
121.   setNames(topTableRUVg$gene_id)
122. pwf_length <- nullp(deVec, bias.data = topTableRUVg$gene_length)
123. ```
124. ```{r}
125. head(pwf_length)
126. ```
127.
128. ```{r}
129. gene2Type <- topTableRUVg %>%
130.   dplyr::select(gene_id, gene_biotype) %>%
131.   split(f = .$gene_id) %>%
132.   lapply(function(x){x$gene_biotype})
133. head(gene2Type)
134. ```
135.
136. ```{r}
137. goseq(pwf_length, gene2cat = gene2Type)
138. ```
139.
140. ```{r}
141. goseq(pwf_length, gene2cat = gene2Type, method = "Hypergeometric")
142. ```
143.
144. # H hallmark gene sets
145. ```{r}
146. Hh <- msigdb(species = "Homo sapiens", category = "H")
147. Hh
148.
149. Hh %>%
150.   group_by(gs_name) %>%
151.   tally()
152.
153. ttx <- read.csv("C:/Users/cronos/Box
  Sync/Desktop/070619/topTableRuvgEMSalmon.csv", row.names=1) %>%
154.   as.data.frame() %>%
155.   dplyr::filter(!is.na(entrezid)) %>%
156.   dplyr::select(gene_id, entrezgene=entrezid)
157. ttx<-transform(ttx, entrezgene = as.integer(entrezgene))
158.
159. Hh <- ttx %>%
160.   left_join(
161.     msigdb(species = "Homo sapiens", category = "H"), by =
162.       c("entrezgene" = "entrez_gene")
163.   ) %>%
164.   dplyr::filter(!is.na(gs_id)) %>%
165.   distinct(gene_id, gs_name, .keep_all = TRUE)
166.
167. ```{r}
168. Hh %>%

```

```

169.   group_by(gs_name) %>%
170.     tally()
171.
172. HhByGene <- Hh %>%
173.   split(f = .$gene_id) %>%
174.   lapply(function(x){x$gs_name})
175. head(HhByGene)
176.
177. goseq(pwf_length, gene2cat = HhByGene)
178. ```
179. ```{r}
180. HhGoseq <- goseq(pwf_length, gene2cat = HhByGene) %>%
181.   as_tibble() %>%
182.   dplyr::select(-under_represented_pvalue) %>%
183.   mutate(adjP = p.adjust(over_represented_pvalue, "bonferroni"))
184. ```
185.
186. ```{r}
187. HhGoseq %>%
188.   dplyr::filter(adjP < 0.05) %>%
189.   dplyr::select(category, nDE = numDEInCat, N = numInCat, p =
190.     over_represented_pvalue, adjP) %>%
191.   pander()
192. ```
193. #GSEA
194. The basic concept which underlies GSEA is that we walk down the
195. ranked list, and the 'enrichment score' increases every time we hit a
196. gene within our gene-set, whilst it decreases every time we don't.
197. The details of the scoring system aren't really relevant, but we look
198. for the extreme enrichment scores within a gene set, and that appear
199. to be more extreme than others when permuting the gene set labels
200. amongst the genes.
201. ```{r}
202. rnkIDs <- topTableRUVg %>%
203.   mutate(rnk = -sign(logFC) * log10(PValue)) %>%
204.   arrange(desc(rnk)) %>%
205.   with(
206.     structure(rnk, names = gene_name)
207.   )
208. ```
209.
210. library(clusterProfiler)
211. HhGsea <- GSEA(
212.   rnkIDs,
213.   nPerm = 1e6,
214.   TERM2GENE = dplyr::select(Hh, term = gs_name, gene =
215.     gene_symbol),
216.   pvalueCutoff = 0.05,
217.   pAdjustMethod = "bonferroni"
218. )
219. ```
220.
221. as_tibble(HhGsea@result)
222. ```
223.
224. ```{r}
225. gseaplot(HhGsea, geneSetID = "HALLMARK_G2M_CHECKPOINT", by =
226.   "runningScore", title = "HALLMARK_G2M_CHECKPOINT")
227. ```

```

```
221.
222. ```{r}
223. library(enrichplot)
224. gseaplot2(HhGsea, geneSetID = "HALLMARK_G2M_CHECKPOINT", title =
  "HALLMARK_G2M_CHECKPOINT")
225. ```
226.
227. ```{r fig.height=10, fig.width=8.5}
228. barcodePlots <- HhGsea@result$ID[1:5] %>%
229.   lapply(function(x) {
230.     gseaplot2(HhGsea, geneSetID = x, base_size = 8, title = x) +
231.     ylim(c(-1, 1)*0.75)
232.   }
233. )
234. cowplot::plot_grid(plotlist = barcodePlots, ncol = 2)
235. ```
236.
```

## REFERENCES

- ABRAHAM, N. G. & KAPPAS, A. 2008. Pharmacological and clinical aspects of heme oxygenase. *Pharmacol Rev*, 60, 79-127.
- ABRAMSON, O., QIU, S. & ERLE, D. J. 2001. Preferential production of interferon- $\gamma$  by CD4<sup>+</sup> T cells expressing the homing receptor integrin  $\alpha 4/\beta 7$ . *Immunology*, 103, 155-163.
- AGGARWAL, S., GHILARDI, N., XIE, M. H., DE SAUVAGE, F. J. & GURNEY, A. L. 2003. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J Biol Chem*, 278, 1910-4.
- AL-AOUKATY, A., ROLSTAD, B., GIAID, A. & MAGHAZACHI, A. A. 1998. MIP-3alpha, MIP-3beta and fractalkine induce the locomotion and the mobilization of intracellular calcium, and activate the heterotrimeric G proteins in human natural killer cells. *Immunology*, 95, 618-24.
- ALLARD, B., LONGHI, M. S., ROBSON, S. C. & STAGG, J. 2017. The ectonucleotidases CD39 and CD73: Novel checkpoint inhibitor targets. *Immunol Rev*, 276, 121-144.
- AMIEL, J., ESPINOSA-PARRILLA, Y., STEFFANN, J., GOSSET, P., PELET, A., PRIEUR, M., BOUTE, O., CHOISSET, A., LACOMBE, D., PHILIP, N., LE MERRER, M., TANAKA, H., TILL, M., TOURAINE, R., TOUTAIN, A., VEKEMANS, M., MUNNICH, A. & LYONNET, S. 2001. Large-scale deletions and SMADIP1 truncating mutations in syndromic Hirschsprung disease with involvement of midline structures. *Am J Hum Genet*, 69, 1370-7.
- ANDREWS, S. 2010. FASTQC. A quality control tool for high throughput sequence data.
- ANNUNZIATO, F., GALLI, G., NAPPI, F., COSMI, L., MANETTI, R., MAGGI, E., ENSOLI, B. & ROMAGNANI, S. 2000. Limited expression of R5-tropic HIV-1 in CCR5-positive type 1-polarized T cells explained by their ability to produce RANTES, MIP-1alpha, and MIP-1beta. *Blood*, 95, 1167-74.
- ANTONIOLI, L., PACHER, P., VIZI, E. S. & HASKÓ, G. 2013. CD39 and CD73 in immunity and inflammation. *Trends Mol Med*, 19, 355-67.
- ARONICA, M. A., MORA, A. L., MITCHELL, D. B., FINN, P. W., JOHNSON, J. E., SHELLER, J. R. & BOOTHBY, M. R. 1999. Preferential role for NF-kappa B/Rel signaling in the type 1 but not type 2 T cell-dependent immune response in vivo. *J Immunol*, 163, 5116-24.
- AVULA, L. R., KNAPEN, D., BUCKINX, R., VERGAUWEN, L., ADRIAENSEN, D., VAN NASSAUW, L. & TIMMERMANS, J.-P. 2012. Whole-genome microarray analysis

- and functional characterization reveal distinct gene expression profiles and patterns in two mouse models of ileal inflammation. *BMC Genomics*, 13, 377.
- BAMIAS, G., KAL TSA, G., SIAKAVELLAS, S. I., PAPAXOINIS, K., ZAMPELI, E., MICHPOULOS, S., ZOUBOULIS-VAFIADIS, I. & LADAS, S. D. 2010. High intestinal and systemic levels of decoy receptor 3 (DcR3) and its ligand TL1A in active ulcerative colitis. *Clinical Immunology*, 137, 242-249.
- BARNES, M. J. & POWRIE, F. 2009. Regulatory T cells reinforce intestinal homeostasis. *Immunity*, 31, 401-11.
- BARRY, S. C., HARDER, B., BRZEZINSKI, M., FLINT, L. Y., SEPPEN, J. & OSBORNE, W. R. 2001. Lentivirus vectors encoding both central polypurine tract and posttranscriptional regulatory element provide enhanced transduction and transgene expression. *Hum Gene Ther*, 12, 1103-8.
- BECHT, E., MCINNES, L., HEALY, J., DUTERTRE, C. A., KWOK, I. W. H., NG, L. G., GINHOUX, F. & NEWELL, E. W. 2018. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*.
- BELNIAK, E., STELMASIAK, Z. & PAPUC, E. 2007. [Multiple sclerosis and other autoimmune diseases]. *Neurol Neurochir Pol*, 41, 259-66.
- BENNETT, C. L., CHRISTIE, J., RAMSDELL, F., BRUNKOW, M. E., FERGUSON, P. J., WHITESELL, L., KELLY, T. E., SAULSBURY, F. T., CHANCE, P. F. & OCHS, H. D. 2001. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet*, 27, 20-1.
- BERNSTEIN, B. E., STAMATOYANNOPOULOS, J. A., COSTELLO, J. F., REN, B., MILOSAVLJEVIC, A., MEISSNER, A., KELLIS, M., MARRA, M. A., BEAUDET, A. L., ECKER, J. R., FARNHAM, P. J., HIRST, M., LANDER, E. S., MIKKELSEN, T. S. & THOMSON, J. A. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*, 28, 1045-1048.
- BEYER, M., THABET, Y., MÜLLER, R.-U., SADLON, T., CLASSEN, S., LAHL, K., BASU, S., ZHOU, X., BAILEY-BUCKTROUT, S. L., KREBS, W., SCHÖNFELD, E. A., BÖTTCHER, J., GOLOVINA, T., MAYER, C. T., HOFMANN, A., SOMMER, D., DEBEY-PASCHER, S., ENDL, E., LIMMER, A., HIPPEN, K. L., BLAZAR, B. R., BALDERAS, R., QUA ST, T., WAHA, A., MAYER, G., FAMULOK, M., KNOLLE, P. A., WICKENHAUSER, C., KOLANUS, W., SCHERMER, B., BLUESTONE, J. A., BARRY, S. C., SPARWASSER, T., RILEY, J. L. & SCHULTZE, J. L. 2011. Repression of SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nature immunology*, 12, 898-907.

- BORGES, T. J., WIETEN, L., VAN HERWIJNEN, M. J., BROERE, F., VAN DER ZEE, R., BONORINO, C. & VAN EDEN, W. 2012. The anti-inflammatory mechanisms of Hsp70. *Front Immunol*, 3, 95.
- BRACKEN, C. P., KHEW-GOODALL, Y. & GOODALL, G. J. 2015. Network-Based Approaches to Understand the Roles of miR-200 and Other microRNAs in Cancer. *Cancer Research*, 75, 2594.
- BROWN, C. Y., DAYAN, S., WONG, S. W., KACZMAREK, A., HOPE, C. M., PEDERSON, S. M., ARNET, V., GOODALL, G. J., RUSSELL, D., SADLON, T. J. & BARRY, S. C. 2018. FOXP3 and miR-155 cooperate to control the invasive potential of human breast cancer cells by down regulating ZEB2 independently of ZEB1. *Oncotarget*, 9, 27708-27727.
- BROWN, C. Y., SADLON, T., GARGETT, T., MELVILLE, E., ZHANG, R., DRABSCH, Y., LING, M., STRATHDEE, C. A., GONDA, T. J. & BARRY, S. C. 2010. Robust, reversible gene knockdown using a single lentiviral short hairpin RNA vector. *Hum Gene Ther*, 21, 1005-17.
- BROWN, D. M. 2010. Cytolytic CD4 cells: Direct mediators in infectious disease and malignancy. *Cellular immunology*, 262, 89-95.
- BRUNKOW, M. E., JEFFERY, E. W., HJERRILD, K. A., PAEPER, B., CLARK, L. B., YASAYKO, S. A., WILKINSON, J. E., GALAS, D., ZIEGLER, S. F. & RAMSDELL, F. 2001. Disruption of a new forkhead/winged-helix protein, scurf, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat Genet*, 27, 68-73.
- CALDERON, D., NGUYEN, M. L. T., MEZGER, A., KATHIRIA, A., MÜLLER, F., NGUYEN, V., LESCANO, N., WU, B., TROMBETTA, J., RIBADO, J. V., KNOWLES, D. A., GAO, Z., BLAESCHKE, F., PARENT, A. V., BURT, T. D., ANDERSON, M. S., CRISWELL, L. A., GREENLEAF, W. J., MARSON, A. & PRITCHARD, J. K. 2019. Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nature Genetics*, 51, 1494-1505.
- CHAUDHRY, A. & RUDENSKY, A. Y. 2013. Control of inflammation by integration of environmental cues by regulatory T cells. *J Clin Invest*, 123, 939-44.
- CLARK, L. B., APPLEBY, M. W., BRUNKOW, M. E., WILKINSON, J. E., ZIEGLER, S. F. & RAMSDELL, F. 1999. Cellular and molecular characterization of the scurfy mouse mutant. *J Immunol*, 162, 2546-54.
- COLLISON, L. W., WORKMAN, C. J., KUO, T. T., BOYD, K., WANG, Y., VIGNALI, K. M., CROSS, R., SEHY, D., BLUMBERG, R. S. & VIGNALI, D. A. 2007. The inhibitory cytokine IL-35 contributes to regulatory T-cell function. *Nature*, 450, 566-9.

- CRAENE, B. D. & BERX, G. 2013. Regulatory networks defining EMT during cancer initiation and progression. *Nature Reviews Cancer*, 13, 97-110.
- CRONKITE, D. A. & STRUTT, T. M. 2018. The Regulation of Inflammation by Innate and Adaptive Lymphocytes. *J Immunol Res*, 2018, 1467538.
- DARDALHON, V., AWASTHI, A., KWON, H., GALILEOS, G., GAO, W., SOBEL, R. A., MITSDOERFFER, M., STROM, T. B., ELYAMAN, W., HO, I. C., KHOURY, S., OUKKA, M. & KUCHROO, V. K. 2008. IL-4 inhibits TGF-beta-induced Foxp3+ T cells and, together with TGF-beta, generates IL-9+ IL-10+ Foxp3(-) effector T cells. *Nat Immunol*, 9, 1347-55.
- DASTOT-LE MOAL, F., WILSON, M., MOWAT, D., COLLOT, N., NIEL, F. & GOOSSENS, M. 2007. ZFHX1B mutations in patients with Mowat-Wilson syndrome. *Hum Mutat*, 28, 313-21.
- DAVIS, C. A., HITZ, B. C., SLOAN, C. A., CHAN, E. T., DAVIDSON, J. M., GABDANK, I., HILTON, J. A., JAIN, K., BAYMURADOV, U. K., NARAYANAN, A. K., ONATE, K. C., GRAHAM, K., MIYASATO, S. R., DRESZER, T. R., STRATTAN, J. S., JOLANKI, O., TANAKA, F. Y. & CHERRY, J. M. 2018. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, 46, D794-d801.
- DE BOER, R. J. & PERELSON, A. S. 2013. Quantifying T lymphocyte turnover. *J Theor Biol*, 327, 45-87.
- DE BOER, R. J., PERELSON, A. S. & RIBEIRO, R. M. 2012. Modelling deuterium labelling of lymphocytes with temporal and/or kinetic heterogeneity. *J R Soc Interface*, 9, 2191-200.
- DE CONINCK, S., BERX, G., TAGHON, T., VAN VLIERBERGHE, P. & GOOSSENS, S. 2019. ZEB2 in T-cells and T-ALL. *Advances in Biological Regulation*, 74, 100639.
- DE JONG, J. M. H., SCHUURHUIS, D. H., IOAN-FACSINAY, A., WELLING, M. M., CAMPS, M. G. M., VAN DER VOORT, E. I. H., HUIZINGA, T. W. J., OSSENDORP, F., VERBEEK, J. S. & TOES, R. E. M. 2006. Dendritic cells, but not macrophages or B cells, activate major histocompatibility complex class II-restricted CD4(+) T cells upon immune-complex uptake in vivo. *Immunology*, 119, 499-506.
- DEAGLIO, S., DWYER, K. M., GAO, W., FRIEDMAN, D., USHEVA, A., ERAT, A., CHEN, J. F., ENJYOJI, K., LINDEN, J., OUKKA, M., KUCHROO, V. K., STROM, T. B. & ROBSON, S. C. 2007. Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J Exp Med*, 204, 1257-65.



- DENG, H. T., LIU, H. L., ZHAI, B. B., ZHANG, K., XU, G. C., PENG, X. M., ZHANG, Q. Z. & LI, L. Y. 2017. Vascular endothelial growth factor suppresses TNFSF15 production in endothelial cells by stimulating miR-31 and miR-20a expression via activation of Akt and Erk signals. *FEBS Open Bio*, 7, 108-117.
- DI GIOVANGIULIO, M., RIZZO, A., FRANZÈ, E., CAPRIOLI, F., FACCIOTTI, F., ONALI, S., FAVALE, A., STOLFI, C., FEHLING, H. J., MONTELEONE, G. & FANTINI, M. C. 2019. Tbet Expression in Regulatory T Cells Is Required to Initiate Th1-Mediated Colitis. *Front Immunol*, 10, 2158.
- DIAZ-RIASCOS, Z. V., GINESTA, M. M., FABREGAT, J., SERRANO, T., BUSQUETS, J., BUSCAIL, L., CORDELIER, P. & CAPELLÁ, G. 2019. Expression and Role of MicroRNAs from the miR-200 Family in the Tumor Formation and Metastatic Propensity of Pancreatic Cancer. *Mol Ther Nucleic Acids*, 17, 491-503.
- DOMINGUEZ-VILLAR, M. & HAFLER, D. A. 2018. Regulatory T cells in autoimmune disease. *Nature Immunology*, 19, 665-673.
- DOMINGUEZ, C. X., AMEZQUITA, R. A., GUAN, T., MARSHALL, H. D., JOSHI, N. S., KLEINSTEIN, S. H. & KAECH, S. M. 2015. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *The Journal of Experimental Medicine*, 212, 2041-2056.
- DUHEN, T., DUHEN, R., LANZAVECCHIA, A., SALLUSTO, F. & CAMPBELL, D. J. 2012. Functionally distinct subsets of human FOXP3+ Treg cells that phenotypically mirror effector Th cells. *Blood*, 119, 4430-40.
- DUREK, P., NORDSTROM, K., GASPARONI, G., SALHAB, A., KRESSLER, C., DE ALMEIDA, M., BASSLER, K., ULAS, T., SCHMIDT, F., XIONG, J., GLAZAR, P., KLIRONOMOS, F., SINHA, A., KINKLEY, S., YANG, X., ARRIGONI, L., AMIRABAD, A. D., ARDAKANI, F. B., FEUERBACH, L., GORKA, O., EBERT, P., MULLER, F., LI, N., FRISCHBUTTER, S., SCHLICKEISER, S., CENDON, C., FROHLER, S., FELDER, B., GASPARONI, N., IMBUSCH, C. D., HUTTER, B., ZIPPRICH, G., TAUCHMANN, Y., REINKE, S., WASSILEW, G., HOFFMANN, U., RICHTER, A. S., SIEVERLING, L., CHANG, H. D., SYRBE, U., KALUS, U., EILS, J., BRORS, B., MANKE, T., RULAND, J., LENGAUER, T., RAJEWSKY, N., CHEN, W., DONG, J., SAWITZKI, B., CHUNG, H. R., ROSENSTIEL, P., SCHULZ, M. H., SCHULTZE, J. L., RADBRUCH, A., WALTER, J., HAMANN, A. & POLANSKY, J. K. 2016. Epigenomic Profiling of Human CD4+ T Cells Supports a Linear Differentiation Model and Highlights Molecular Regulators of Memory Development. *Immunity*, 45, 1148-1161.
- EGGENHUIZEN, P. J., NG, B. H. & OOI, J. D. 2020. Treg Enhancing Therapies to Treat Autoimmune Diseases. *Int J Mol Sci*, 21.

- EIZ-VESPER, B. & SCHMETZER, H. M. 2020. Antigen-Presenting Cells: Potential of Proven und New Players in Immune Therapies. *Transfusion Medicine and Hemotherapy*, 47, 429-431.
- ESCOBAR, T. M., KANELLOPOULOU, C., KUGLER, D. G., KILARU, G., NGUYEN, C. K., NAGARAJAN, V., BHAIKAVABHOTLA, R. K., NORTHRUP, D., ZAHR, R., BURR, P., LIU, X., ZHAO, K., SHER, A., JANKOVIC, D., ZHU, J. & MULJO, S. A. 2014. miR-155 activates cytokine gene expression in Th17 cells by regulating the DNA-binding protein Jarid2 to relieve polycomb-mediated repression. *Immunity*, 40, 865-79.
- EYERICH, S., EYERICH, K., PENNINO, D., CARBONE, T., NASORRI, F., PALLOTTA, S., CIANFARANI, F., ODORISIO, T., TRAILD-HOFFMANN, C., BEHRENDT, H., DURHAM, S. R., SCHMIDT-WEBER, C. B. & CAVANI, A. 2009. Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling. *J Clin Invest*, 119, 3573-85.
- FAHLÉN, L., READ, S., GORELIK, L., HURST, S. D., COFFMAN, R. L., FLAVELL, R. A. & POWRIE, F. 2005. T cells that cannot respond to TGF-beta escape control by CD4(+)CD25(+) regulatory T cells. *J Exp Med*, 201, 737-46.
- FARDI, M., ALIVAND, M., BARADARAN, B., FARSHDOUSTI HAGH, M. & SOLALI, S. 2019. The crucial role of ZEB2: From development to epithelial-to-mesenchymal transition and cancer complexity. *J Cell Physiol*.
- FENG, T., CAO, A. T., WEAVER, C. T., ELSON, C. O. & CONG, Y. 2011. Interleukin-12 converts Foxp3+ regulatory T cells to interferon- $\gamma$ -producing Foxp3+ T cells that inhibit colitis. *Gastroenterology*, 140, 2031-43.
- FESKE, S., PRAKRIYA, M., RAO, A. & LEWIS, R. S. 2005. A severe defect in CRAC Ca<sup>2+</sup> channel activation and altered K<sup>+</sup> channel gating in T cells from immunodeficient patients. *J Exp Med*, 202, 651-62.
- FLETCHER, J. M., LONERGAN, R., COSTELLOE, L., KINSELLA, K., MORAN, B., O'FARRELLY, C., TUBRIDY, N. & MILLS, K. H. 2009. CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. *J Immunol*, 183, 7602-10.
- FLOESS, S., FREYER, J., SIEWERT, C., BARON, U., OLEK, S., POLANSKY, J., SCHLAWKE, K., CHANG, H.-D., BOPP, T., SCHMITT, E., KLEIN-HESSLING, S., SERFLING, E., HAMANN, A. & HUEHN, J. 2007. Epigenetic Control of the foxp3 Locus in Regulatory T Cells. *PLOS Biology*, 5, e38.
- FRANKISH, A., DIEKHANS, M., FERREIRA, A. M., JOHNSON, R., JUNGREIS, I., LOVELAND, J., MUDGE, J. M., SISU, C., WRIGHT, J., ARMSTRONG, J.,

- BARNES, I., BERRY, A., BIGNELL, A., CARBONELL SALA, S., CHRAST, J., CUNNINGHAM, F., DI DOMENICO, T., DONALDSON, S., FIDDES, I. T., GARCÍA GIRÓN, C., GONZALEZ, J. M., GREGO, T., HARDY, M., HOURLIER, T., HUNT, T., IZUOGU, O. G., LAGARDE, J., MARTIN, F. J., MARTÍNEZ, L., MOHANAN, S., MUIR, P., NAVARRO, F. C. P., PARKER, A., PEI, B., POZO, F., RUFFIER, M., SCHMITT, B. M., STAPLETON, E., SUNER, M. M., SYCHEVA, I., USZCZYNSKA-RATAJCZAK, B., XU, J., YATES, A., ZERBINO, D., ZHANG, Y., AKEN, B., CHOUDHARY, J. S., GERSTEIN, M., GUIGÓ, R., HUBBARD, T. J. P., KELLIS, M., PATEN, B., REYMOND, A., TRESS, M. L. & FLICEK, P. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*, 47, D766-d773.
- FREDENBURGH, L. E., MERZ, A. A. & CHENG, S. 2015. Haeme oxygenase signalling pathway: implications for cardiovascular disease. *Eur Heart J*, 36, 1512-8.
- FREELEY, S., CARDONE, J., GÜNTHER, S. C., WEST, E. E., REINHECKEL, T., WATTS, C., KEMPER, C. & KOLEV, M. V. 2018. Asparaginyl Endopeptidase (Legumain) Supports Human Th1 Induction via Cathepsin L-Mediated Intracellular C3 Activation. *Frontiers in immunology*, 9, 2449-2449.
- FU, J., WANG, D., YU, Y., HEINRICHS, J., WU, Y., SCHUTT, S., KAOSAARD, K., LIU, C., HAARBERG, K., BASTIAN, D., MCDONALD, D. G., ANASETTI, C. & YU, X.-Z. 2015. T-bet is Critical for the Development of Acute Graft-versus-Host Disease through Controlling T-Cell Differentiation and Function. *Journal of immunology (Baltimore, Md. : 1950)*, 194, 388-397.
- FUCHIZAWA, T., ADACHI, Y., ITO, Y., HIGASHIYAMA, H., KANEGANE, H., FUTATANI, T., KOBAYASHI, I., KAMACHI, Y., SAKAMOTO, T., TSUGE, I., TANAKA, H., BANHAM, A. H., OCHS, H. D. & MIYAWAKI, T. 2007. Developmental changes of FOXP3-expressing CD4+CD25+ regulatory T cells and their impairment in patients with FOXP3 gene mutations. *Clin Immunol*, 125, 237-46.
- GAGNON, J. D. & ANSEL, K. M. 2019. MicroRNA regulation of CD8(+) T cell responses. *Noncoding RNA Investig*, 3.
- GALLI, F., AGUILERA, J. V., PALERMO, B., MARKOVIC, S. N., NISTICÒ, P. & SIGNORE, A. 2020. Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy. *Journal of Experimental & Clinical Cancer Research*, 39, 89.
- GÁLVEZ, J., GARRIDO, M., MERLOS, M., TORRES, M. I. & ZARZUELO, A. 2000. Intestinal anti-inflammatory activity of UR-12746, a novel 5-ASA conjugate, on acute and chronic experimental colitis in the rat. *Br J Pharmacol*, 130, 1949-59.

- EGINAT, J., PARONI, M., MAGLIE, S., ALFEN, J. S., KASTIRR, I., GRUARIN, P., DE SIMONE, M., PAGANI, M. & ABRIGNANI, S. 2014. Plasticity of human CD4 T cell subsets. *Front Immunol*, 5, 630.
- GLADKA, M., DE LEEUW, A., KOHELA, A., MOLENAAR, B., VERSTEEG, D., KOOIJMAN, L., VAN GELDORP, M. & VAN ROOIJ, E. 2020. ZEB2 regulates a transcriptional network of calcium-handling genes in the injured heart. *European Heart Journal*, 41.
- GONDEK, D. C., LU, L. F., QUEZADA, S. A., SAKAGUCHI, S. & NOELLE, R. J. 2005. Cutting edge: contact-mediated suppression by CD4+CD25+ regulatory cells involves a granzyme B-dependent, perforin-independent mechanism. *J Immunol*, 174, 1783-6.
- GONZALEZ, S., GONZÁLEZ-RODRÍGUEZ, A. P., SUÁREZ-ÁLVAREZ, B., LÓPEZ-SOTO, A., HUERGO-ZAPICO, L. & LOPEZ-LARREA, C. 2011. Conceptual aspects of self and nonself discrimination. *Self Nonself*, 2, 19-25.
- GOOSSENS, S., JANZEN, V., BARTUNKOVA, S., YOKOMIZO, T., DROGAT, B., CRISAN, M., HAIGH, K., SEUNTJENS, E., UMANS, L., RIEDT, T., BOGAERT, P., HAENEBALCKE, L., BERX, G., DZIERZAK, E., HUYLEBROECK, D. & HAIGH, J. J. 2011. The EMT regulator Zeb2/Sip1 is essential for murine embryonic hematopoietic stem/progenitor cell differentiation and mobilization. *Blood*, 117, 5620-30.
- GOOSSENS, S., RADAELLI, E., BLANCHET, O., DURINCK, K., VAN DER MEULEN, J., PEIRS, S., TAGHON, T., TREMBLAY, C. S., COSTA, M., FARHANG GHahremani, M., DE MEDTS, J., BARTUNKOVA, S., HAIGH, K., SCHWAB, C., FARLA, N., PIETERS, T., MATTHIJSENS, F., VAN ROY, N., BEST, J. A., DESWARTE, K., BOGAERT, P., CARMICHAEL, C., RICKARD, A., SURYANI, S., BRACKEN, L. S., ALSERIHI, R., CANTE-BARRETT, K., HAENEBALCKE, L., CLAPPIER, E., RONDOU, P., SLOWICKA, K., HUYLEBROECK, D., GOLDRATH, A. W., JANZEN, V., MCCORMACK, M. P., LOCK, R. B., CURTIS, D. J., HARRISON, C., BERX, G., SPELEMAN, F., MEIJERINK, J. P., SOULIER, J., VAN VLIERBERGHE, P. & HAIGH, J. J. 2015. ZEB2 drives immature T-cell lymphoblastic leukaemia development via enhanced tumour-initiating potential and IL-7 receptor signalling. *Nat Commun*, 6, 5794.
- GOOSSENS, S., WANG, J., TREMBLAY, C. S., DE MEDTS, J., T'SAS, S., NGUYEN, T., SAW, J., HAIGH, K., CURTIS, D. J., VAN VLIERBERGHE, P., BERX, G., TAGHON, T. & HAIGH, J. J. 2019. ZEB2 and LMO2 drive immature T-cell lymphoblastic leukemia via distinct oncogenic mechanisms. *Haematologica*, 104, 1608-1616.

- GRAINGE, M. J., WEST, J. & CARD, T. R. 2010. Venous thromboembolism during active disease and remission in inflammatory bowel disease: a cohort study. *Lancet*, 375, 657-63.
- GREGORY, P. A., BERT, A. G., PATERSON, E. L., BARRY, S. C., TSYKIN, A., FARSHID, G., VADAS, M. A., KHEW-GOODALL, Y. & GOODALL, G. J. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol*, 10, 593-601.
- GRICKS, C. S. & GRIBBEN, J. G. 2003. Cytotoxic T cell responses against immunoglobulin in malignant and normal B cells: implications for tumor immunity and autoimmunity. *Curr Pharm Des*, 9, 1889-903.
- GROSSMAN, W. J., VERBSKY, J. W., TOLLEFSEN, B. L., KEMPER, C., ATKINSON, J. P. & LEY, T. J. 2004. Differential expression of granzymes A and B in human cytotoxic lymphocyte subsets and T regulatory cells. *Blood*, 104, 2840-8.
- HA, T.-Y. 2011. The Role of MicroRNAs in Regulatory T Cells and in the Immune Response. *Immune Network*, 11, 11-41.
- HALL, A. O., BEITING, D. P., TATO, C., JOHN, B., OLDENHOVE, G., LOMBANA, C. G., PRITCHARD, G. H., SILVER, J. S., BOULADOUX, N., STUMHOFER, J. S., HARRIS, T. H., GRAINGER, J., WOJNO, E. D., WAGAGE, S., ROOS, D. S., SCOTT, P., TURKA, L. A., CHERRY, S., REINER, S. L., CUA, D., BELKAID, Y., ELLOSO, M. M. & HUNTER, C. A. 2012. The cytokines interleukin 27 and interferon-gamma promote distinct Treg cell populations required to limit infection-induced pathology. *Immunity*, 37, 511-23.
- HARA, M., KINGSLEY, C. I., NIIMI, M., READ, S., TURVEY, S. E., BUSHELL, A. R., MORRIS, P. J., POWRIE, F. & WOOD, K. J. 2001. IL-10 is required for regulatory T cells to mediate tolerance to alloantigens in vivo. *J Immunol*, 166, 3789-96.
- HARIBHAI, D., WILLIAMS, J. B., JIA, S., NICKERSON, D., SCHMITT, E. G., EDWARDS, B., ZIEGELBAUER, J., YASSAI, M., LI, S. H., RELLAND, L. M., WISE, P. M., CHEN, A., ZHENG, Y. Q., SIMPSON, P. M., GORSKI, J., SALZMAN, N. H., HESSNER, M. J., CHATILA, T. A. & WILLIAMS, C. B. 2011. A requisite role for induced regulatory T cells in tolerance based on expanding antigen receptor diversity. *Immunity*, 35, 109-22.
- HENDEL, A., BAK, R. O., CLARK, J. T., KENNEDY, A. B., RYAN, D. E., ROY, S., STEINFELD, I., LUNSTAD, B. D., KAISER, R. J., WILKENS, A. B., BACCHETTA, R., TSALENKO, A., DELLINGER, D., BRUHN, L. & PORTEUS, M. H. 2015. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology*, 33, 985-989.

- HERTWECK, A., EVANS, CATHERINE M., ESKANDARPOUR, M., LAU, JONATHAN C. H., OLEINIKA, K., JACKSON, I., KELLY, A., AMBROSE, J., ADAMSON, P., COUSINS, DAVID J., LAVENDER, P., CALDER, VIRGINIA L., LORD, GRAHAM M. & JENNER, RICHARD G. 2016. T-bet Activates Th1 Genes through Mediator and the Super Elongation Complex. *Cell Reports*, 15, 2756-2770.
- HO, I. C., TAI, T.-S. & PAI, S.-Y. 2009. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nature reviews. Immunology*, 9, 125-135.
- HÖLLBACHER, B., DUHEN, T., MOTLEY, S., KLICZNIK, M. M., GRATZ, I. K. & CAMPBELL, D. J. 2020. Transcriptomic Profiling of Human Effector and Regulatory T Cell Subsets Identifies Predictive Population Signatures. *Immunohorizons*, 4, 585-596.
- HOMMES, D. W., MIKHAILOVA, T. L., STOINOV, S., STIMAC, D., VUCELIC, B., LONOVICS, J., ZÁKUCIOVÁ, M., D'HAENS, G., VAN ASSCHE, G., BA, S., LEE, S. & PEARCE, T. 2006. Fontolizumab, a humanised anti-interferon gamma antibody, demonstrates safety and clinical activity in patients with moderate to severe Crohn's disease. *Gut*, 55, 1131-7.
- HOPE, C. M., WELCH, J., MOHANDAS, A., PEDERSON, S., HILL, D., GUNDSAMBUU, B., EASTAFF-LEUNG, N., GROSSE, R., BRESATZ, S., ANG, G., PAPADEMETRIOS, M., ZOLA, H., DUHEN, T., CAMPBELL, D., BROWN, C. Y., KRUMBIEGEL, D., SADLON, T., COUPER, J. J. & BARRY, S. C. 2019. Peptidase inhibitor 16 identifies a human regulatory T-cell subset with reduced FOXP3 expression over the first year of recent onset type 1 diabetes. *Eur J Immunol*, 49, 1235-1250.
- HUANG, H., SIKORA, M. J., ISLAM, S., CHOWDHURY, R. R., CHIEN, Y. H., SCRIBA, T. J., DAVIS, M. M. & STEINMETZ, L. M. 2019. Select sequencing of clonally expanded CD8(+) T cells reveals limits to clonal expansion. *Proc Natl Acad Sci U S A*, 116, 8995-9001.
- HUANG, W., LOGANANTHARAJ, R., SCHROEDER, B., FARGO, D. & LI, L. 2013. PAVIS: a tool for Peak Annotation and Visualization. *Bioinformatics*, 29, 3097-9.
- IMAI, T., HIESHIMA, K., HASKELL, C., BABA, M., NAGIRA, M., NISHIMURA, M., KAKIZAKI, M., TAKAGI, S., NOMIYAMA, H., SCHALL, T. J. & YOSHIE, O. 1997. Identification and molecular characterization of fractalkine receptor CX3CR1, which mediates both leukocyte migration and adhesion. *Cell*, 91, 521-30.
- IRIGUCHI, S., KIKUCHI, N., KANEKO, S., NOGUCHI, E., MORISHIMA, Y., MATSUYAMA, M., YOH, K., TAKAHASHI, S., NAKAUCHI, H. & ISHII, Y. 2015.

- T-cell–restricted T-bet overexpression induces aberrant hematopoiesis of myeloid cells and impairs function of macrophages in the lung. *Blood*, 125, 370-382.
- JANKOVIC, D., KUGLER, D. G. & SHER, A. 2010. IL-10 production by CD4+ effector T cells: a mechanism for self-regulation. *Mucosal immunology*, 3, 239-246.
- JIANG, H. & CHESS, L. 2009. How the immune system achieves self-nonsel self discrimination during adaptive immunity. *Adv Immunol*, 102, 95-133.
- JIN, S., CHIN, J., SEEGER, S., NIEWOEHRER, J., WEISER, B., BEAUCAMP, N., WOODS, J., MURPHY, C., FANNING, A., SHANAHAN, F., NALLY, K., KAJEKAR, R., SALAS, A., PLANELL, N., LOZANO, J., PANES, J., PARMAR, H., DEMARTINO, J., NARULA, S. & THOMAS-KARYAT, D. A. 2013. TL1A/TNFSF15 directly induces proinflammatory cytokines, including TNF $\alpha$ , from CD3+CD161+ T cells to exacerbate gut inflammation. *Mucosal Immunol*, 6, 886-99.
- KELLIE, S. & AL-MANSOUR, Z. 2017. Chapter Four - Overview of the Immune System. In: SKWARCZYNSKI, M. & TOTH, I. (eds.) *Micro and Nanotechnology in Vaccine Development*. William Andrew Publishing.
- KESSELS, H. W. H. G., SCHEPERS, K., VAN DEN BOOM, M. D., TOPHAM, D. J. & SCHUMACHER, T. N. M. 2006. Generation of T Cell Help through a MHC Class I-Restricted TCR. *The Journal of Immunology*, 177, 976.
- KIM, J., HOPE, C. M., PERKINS, G. B., STEAD, S. O., SCAFFIDI, J. C., KETTE, F. D., CARROLL, R. P., BARRY, S. C. & COATES, P. T. 2020. Rapamycin and abundant TCR stimulation are required for the generation of stable human induced regulatory T cells. *Clinical & Translational Immunology*, 9, e1223.
- KOCH, M. A., TUCKER-HEARD, G. S., PERDUE, N. R., KILLEBREW, J. R., URDAHL, K. B. & CAMPBELL, D. J. 2009. T-bet controls regulatory T cell homeostasis and function during type-1 inflammation. *Nature immunology*, 10, 595-602.
- KRETSCHMER, L., FLOSSDORF, M., MIR, J., CHO, Y.-L., PLAMBECK, M., TREISE, I., TOSKA, A., HEINZEL, S., SCHIEMANN, M., BUSCH, D. H. & BUCHHOLZ, V. R. 2020. Differential expansion of T central memory precursor and effector subsets is regulated by division speed. *Nature Communications*, 11, 113.
- KRUEGER, P. D., GOLDBERG, M. F., HONG, S. W., OSUM, K. C., LANGLOIS, R. A., KOTOV, D. I., DILEEPAN, T. & JENKINS, M. K. 2021. Two sequential activation modules control the differentiation of protective T helper-1 (Th1) cells. *Immunity*, 54, 687-701.e4.

- KUANG, D. M., XIAO, X., ZHAO, Q., CHEN, M. M., LI, X. F., LIU, R. X., WEI, Y., OUYANG, F. Z., CHEN, D. P., WU, Y., LAO, X. M., DENG, H. & ZHENG, L. 2014. B7-H1-expressing antigen-presenting cells mediate polarization of protumorigenic Th22 subsets. *J Clin Invest*, 124, 4657-67.
- KULESHOV, M. V., JONES, M. R., ROUILLARD, A. D., FERNANDEZ, N. F., DUAN, Q., WANG, Z., KOPLEV, S., JENKINS, S. L., JAGODNIK, K. M., LACHMANN, A., MCDERMOTT, M. G., MONTEIRO, C. D., GUNDERSEN, G. W. & MA'AYAN, A. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, 44, W90-7.
- LAGOU, V., GARCIA-PEREZ, J. E., SMETS, I., VAN HOREBEEK, L., VANDEBERGH, M., CHEN, L., MALLANTS, K., PREZZEMOLO, T., HILVEN, K., HUMBLET-BARON, S., MOISSE, M., VAN DAMME, P., BOECKXSTAENS, G., BOWNESS, P., DUBOIS, B., DOOLEY, J., LISTON, A. & GORIS, A. 2018. Genetic Architecture of Adaptive Immune System Identifies Key Immune Regulators. *Cell Rep*, 25, 798-810.e6.
- LANTELME, E., MANTOVANI, S., PALERMO, B., CAMPANELLI, R., SALLUSTO, F. & GIACHINO, C. 2001. Kinetics of GATA-3 gene expression in early polarizing and committed human T cells. *Immunology*, 102, 123-30.
- LE DEIST, F., HIVROZ, C., PARTISETI, M., THOMAS, C., BUC, H. A., OLEASTRO, M., BELOHRADSKY, B., CHOQUET, D. & FISCHER, A. 1995. A primary T-cell immunodeficiency associated with defective transmembrane calcium influx. *Blood*, 85, 1053-62.
- LEUNG, S., LIU, X., FANG, L., CHEN, X., GUO, T. & ZHANG, J. 2010. The cytokine milieu in the interplay of pathogenic Th1/Th17 cells and regulatory T cells in autoimmune disease. *Cellular & molecular immunology*, 7, 182-189.
- LEVINE, A. G., MENDOZA, A., HEMMERS, S., MOLTEDO, B., NIEC, R. E., SCHIZAS, M., HOYOS, B. E., PUTINTSEVA, E. V., CHAUDHRY, A., DIKIY, S., FUJISAWA, S., CHUDAKOV, D. M., TREUTING, P. M. & RUDENSKY, A. Y. 2017. Stability and function of regulatory T cells expressing the transcription factor T-bet. *Nature*, 546, 421-425.
- LI, H., MAR, B. G., ZHANG, H., PURAM, R. V., VAZQUEZ, F., WEIR, B. A., HAHN, W. C., EBERT, B. & PELLMAN, D. 2016a. The EMT regulator ZEB2 is a novel dependency of human and murine acute myeloid leukemia. *Blood*.
- LI, J., RIEDT, T., GOOSSENS, S., CARRILLO GARCIA, C., SZCZEPANSKI, S., BRANDES, M., PIETERS, T., DOBROSCH, L., GUTGEMANN, I., FARLA, N., RADAELLI, E., HULPIAU, P., MALLELA, N., FROHLICH, H., LA STARZA, R.,



- MATTEUCCI, C., CHEN, T., BROSSART, P., MECUCCI, C., HUYLEBROECK, D., HAIGH, J. J. & JANZEN, V. 2016b. The EMT transcription factor *Zeb2* controls adult murine hematopoietic differentiation by regulating cytokine signaling. *Blood*.
- LINDGREEN, S. 2012. AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC research notes*, 5, 337.
- LU, L., MA, J., LI, Z., LAN, Q., CHEN, M., LIU, Y., XIA, Z., WANG, J., HAN, Y., SHI, W., QUESNIAUX, V., RYFFEL, B., BRAND, D., LI, B., LIU, Z. & ZHENG, S. G. 2011. All-Trans Retinoic Acid Promotes TGF- $\beta$ -Induced Tregs via Histone Modification but Not DNA Demethylation on *Foxp3* Gene Locus. *PLoS ONE*, 6, e24590.
- MAHNKE, Y. D., BRODIE, T. M., SALLUSTO, F., ROEDERER, M. & LUGLI, E. 2013. The who's who of T-cell differentiation: Human memory T-cell subsets. *European Journal of Immunology*, 43, 2797-2809.
- MAILLOUX, A. W. & EPLING-BURNETTE, P. K. 2013. Effector memory regulatory T-cell expansion marks a pivotal point of immune escape in myelodysplastic syndromes. *Oncoimmunology*, 2, e22654-e22654.
- MANETTI, R., PARRONCHI, P., GIUDIZI, M. G., PICCINNI, M. P., MAGGI, E., TRINCHIERI, G. & ROMAGNANI, S. 1993. Natural killer cell stimulatory factor (interleukin 12 [IL-12]) induces T helper type 1 (Th1)-specific immune responses and inhibits the development of IL-4-producing Th cells. *J Exp Med*, 177, 1199-204.
- MARSHALL, J. S., WARRINGTON, R., WATSON, W. & KIM, H. L. 2018. An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, 14, 49.
- MARTIN, M. D. & BADOVINAC, V. P. 2018. Defining Memory CD8 T Cell. *Frontiers in Immunology*, 9.
- MCINNES, N., SADLON, T. J., BROWN, C. Y., PEDERSON, S., BEYER, M., SCHULTZE, J. L., MCCOLL, S., GOODALL, G. J. & BARRY, S. C. 2012. FOXP3 and FOXP3-regulated microRNAs suppress SATB1 in breast cancer cells. *Oncogene*, 31, 1045-54.
- MECKIFF, B. J., RAMÍREZ-SUÁSTEGUI, C., FAJARDO, V., CHEE, S. J., KUSNADI, A., SIMON, H., GRIFONI, A., PELOSI, E., WEISKOPF, D., SETTE, A., AY, F., SEUMOIS, G., OTTENSMEIER, C. & VIJAYANAND, P. 2020. Single-Cell Transcriptomic Analysis of SARS-CoV-2 Reactive CD4 (+) T Cells. *Ssrn*, 3641939.

- MIKAMI, N., KAWAKAMI, R., CHEN, K. Y., SUGIMOTO, A., OHKURA, N. & SAKAGUCHI, S. 2020. Epigenetic conversion of conventional T cells into regulatory T cells by CD28 signal deprivation. *Proc Natl Acad Sci U S A*, 117, 12258-12268.
- MONACO, G., CHEN, H., POIDINGER, M., CHEN, J., DE MAGALHÃES, J. P. & LARBI, A. 2016. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*, 32, 2473-80.
- MOSMANN, T. R., CHERWINSKI, H., BOND, M. W., GIEDLIN, M. A. & COFFMAN, R. L. 1986. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *J Immunol*, 136, 2348-57.
- MOWAT, D. R., WILSON, M. J. & GOOSSENS, M. 2003. Mowat-Wilson syndrome. *J Med Genet*, 40, 305-10.
- MULLER, P. Y., JANOVJAK, H., MISEREZ, A. R. & DOBBIE, Z. 2002. Processing of gene expression data generated by quantitative real-time RT-PCR. *Biotechniques*, 32, 1372-4, 1376, 1378-9.
- MURANSKI, P., BORMAN, Z. A., KERKAR, S. P., KLEBANOFF, C. A., JI, Y., SANCHEZ-PEREZ, L., SUKUMAR, M., REGER, R. N., YU, Z., KERN, S. J., ROYCHOUDHURI, R., FERREYRA, G. A., SHEN, W., DURUM, S. K., FEIGENBAUM, L., PALMER, D. C., ANTONY, P. A., CHAN, C. C., LAURENCE, A., DANNER, R. L., GATTINONI, L. & RESTIFO, N. P. 2011. Th17 cells are long lived and retain a stem cell-like molecular signature. *Immunity*, 35, 972-85.
- NGUYEN, G. C., BERNSTEIN, C. N., BITTON, A., CHAN, A. K., GRIFFITHS, A. M., LEONTIADIS, G. I., GEERTS, W., BRESSLER, B., BUTZNER, J. D., CARRIER, M., CHANDE, N., MARSHALL, J. K., WILLIAMS, C. & KEARON, C. 2014. Consensus statements on the risk, prevention, and treatment of venous thromboembolism in inflammatory bowel disease: Canadian Association of Gastroenterology. *Gastroenterology*, 146, 835-848.e6.
- NICOLET, B. P., GUISLAIN, A., VAN ALPHEN, F. P. J., GOMEZ-EERLAND, R., SCHUMACHER, T. N. M., VAN DEN BIGGELAAR, M. & WOLKERS, M. C. 2020. CD29 identifies IFN- $\gamma$ -producing human CD8<sup>+</sup> T cells with an increased cytotoxic potential. *Proceedings of the National Academy of Sciences*, 117, 6686.
- O'SHEA, J. J. & PAUL, W. E. 2010. Mechanisms underlying lineage commitment and plasticity of helper CD4<sup>+</sup> T cells. *Science*, 327, 1098-102.
- OHKURA, N., HAMAGUCHI, M., MORIKAWA, H., SUGIMURA, K., TANAKA, A., ITO, Y., OSAKI, M., TANAKA, Y., YAMASHITA, R., NAKANO, N., HUEHN, J.,

- FEHLING, H. J., SPARWASSER, T., NAKAI, K. & SAKAGUCHI, S. 2012. T cell receptor stimulation-induced epigenetic changes and Foxp3 expression are independent and complementary events required for Treg cell development. *Immunity*, 37, 785-99.
- OLSEN, I., LUNDIN, K. E. & SOLLID, L. M. 2013. Increased frequency of intestinal CD4+ T cells reactive with mycobacteria in patients with Crohn's disease. *Scand J Gastroenterol*, 48, 1278-85.
- OMILUSIK, K. D., BEST, J. A., YU, B., GOOSSENS, S., WEIDEMANN, A., NGUYEN, J. V., SEUNTJENS, E., STRYJEWSKA, A., ZWEIER, C., ROYCHOUDHURI, R., GATTINONI, L., BIRD, L. M., HIGASHI, Y., KONDOH, H., HUYLEBROECK, D., HAIGH, J. & GOLDRATH, A. W. 2015. Transcriptional repressor ZEB2 promotes terminal differentiation of CD8+ effector and memory T cell populations during infection. *J Exp Med*, 212, 2027-39.
- OPSTELTEN, R., DE KIVIT, S., SLOT, M. C., VAN DEN BIGGELAAR, M., IWASZKIEWICZ-GRZEŚ, D., GLIWŃSKI, M., SCOTT, A. M., BLOM, B., TRZONKOWSKI, P., BORST, J., CUADRADO, E. & AMSEN, D. 2020. GPA33: A Marker to Identify Stable Human Regulatory T Cells. *J Immunol*, 204, 3139-3148.
- PANÉS, J. & GRANGER, D. N. 1998. Leukocyte-endothelial cell interactions: molecular mechanisms and implications in gastrointestinal disease. *Gastroenterology*, 114, 1066-90.
- PARK, S.-M., GAUR, A. B., LENGYEL, E. & PETER, M. E. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & development*, 22, 894-907.
- PARKER, D. C. 1993. T cell-dependent B cell activation. *Annu Rev Immunol*, 11, 331-60.
- PARTISETI, M., LE DEIST, F., HIVROZ, C., FISCHER, A., KORN, H. & CHOQUET, D. 1994. The calcium current activated by T cell receptor and store depletion in human lymphocytes is absent in a primary immunodeficiency. *J Biol Chem*, 269, 32327-35.
- PATIL, V. S., MADRIGAL, A., SCHMIEDEL, B. J., CLARKE, J., O'ROURKE, P., DE SILVA, A. D., HARRIS, E., PETERS, B., SEUMOIS, G., WEISKOPF, D., SETTE, A. & VIJAYANAND, P. 2018. Precursors of human CD4(+) cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol*, 3.
- PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 14, 417-419.

- PENG, S. L., SZABO, S. J. & GLIMCHER, L. H. 2002. T-bet regulates IgG class switching and pathogenic. *Proc Natl Acad Sci U S A*, 99, 5545-50.
- PLANK, M. W., KAIKO, G. E., MALTBY, S., WEAVER, J., TAY, H. L., SHEN, W., WILSON, M. S., DURUM, S. K. & FOSTER, P. S. 2017. Th22 Cells Form a Distinct Th Lineage from Th17 Cells In Vitro with Unique Transcriptional Properties and Tbet-Dependent Th1 Plasticity. *J Immunol*, 198, 2182-2190.
- POSTIGO, A. A. & DEAN, D. C. 2000. Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors. *Proc Natl Acad Sci U S A*, 97, 6391-6.
- POWRIE, F., CARLINO, J., LEACH, M. W., MAUZE, S. & COFFMAN, R. L. 1996. A critical role for transforming growth factor-beta but not interleukin 4 in the suppression of T helper type 1-mediated colitis by CD45RB(low) CD4+ T cells. *J Exp Med*, 183, 2669-74.
- PREHN, J. L., MEHDIZADEH, S., LANDERS, C. J., LUO, X., CHA, S. C., WEI, P. & TARGAN, S. R. 2004. Potential role for TL1A, the new TNF-family member and potent costimulator of IFN-gamma, in mucosal inflammation. *Clin Immunol*, 112, 66-77.
- QI, S., SONG, Y., PENG, Y., WANG, H., LONG, H., YU, X., LI, Z., FANG, L., WU, A., LUO, W., ZHEN, Y., ZHOU, Y., CHEN, Y., MAI, C., LIU, Z. & FANG, W. 2012. ZEB2 mediates multiple pathways regulating cell proliferation, migration, invasion, and apoptosis in glioma. *PloS one*, 7, e38842-e38842.
- QURESHI, O. S., ZHENG, Y., NAKAMURA, K., ATTRIDGE, K., MANZOTTI, C., SCHMIDT, E. M., BAKER, J., JEFFERY, L. E., KAUR, S., BRIGGS, Z., HOU, T. Z., FUTTER, C. E., ANDERSON, G., WALKER, L. S. & SANSOM, D. M. 2011. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. *Science*, 332, 600-3.
- R CORE TEAM 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- READ, S., MALMSTRÖM, V. & POWRIE, F. 2000. Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation. *J Exp Med*, 192, 295-302.
- REINISCH, W., DE VILLIERS, W., BENE, L., SIMON, L., RÁCZ, I., KATZ, S., ALTORJAY, I., FEAGAN, B., RIFF, D., BERNSTEIN, C. N., HOMMES, D., RUTGEERTS, P., CORTOT, A., GASPARI, M., CHENG, M., PEARCE, T. & SANDS, B. E. 2010. Fontolizumab in moderate to severe Crohn's disease: a phase 2,

- randomized, double-blind, placebo-controlled, multiple-dose study. *Inflamm Bowel Dis*, 16, 233-42.
- RISSE, D., NGAI, J., SPEED, T. P. & DUDOIT, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*, 32, 896-902.
- ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-40.
- RUDENSKY, A. Y. 2011. Regulatory T Cells and Foxp3. *Immunological reviews*, 241, 260-268.
- RUDRA, D., DEROOS, P., CHAUDHRY, A., NIEC, R. E., ARVEY, A., SAMSTEIN, R. M., LESLIE, C., SHAFFER, S. A., GOODLETT, D. R. & RUDENSKY, A. Y. 2012. Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat Immunol*, 13, 1010-9.
- RUPP, L. J., SCHUMANN, K., ROYBAL, K. T., GATE, R. E., YE, C. J., LIM, W. A. & MARSON, A. 2017. CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Scientific Reports*, 7, 737.
- RUSO, P. S. T., FERREIRA, G. R., CARDOZO, L. E., BÜRGER, M. C., ARIAS-CARRASCO, R., MARUYAMA, S. R., HIRATA, T. D. C., LIMA, D. S., PASSOS, F. M., FUKUTANI, K. F., LEVER, M., SILVA, J. S., MARACAJA-COUTINHO, V. & NAKAYA, H. I. 2018. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinformatics*, 19, 56.
- RYTER, S. W., ALAM, J. & CHOI, A. M. 2006. Heme oxygenase-1/carbon monoxide: from basic science to therapeutic applications. *Physiol Rev*, 86, 583-650.
- SADLON, T. J., WILKINSON, B. G., PEDERSON, S., BROWN, C. Y., BRESATZ, S., GARGETT, T., MELVILLE, E. L., PENG, K., D'ANDREA, R. J., GLONEK, G. G., GOODALL, G. J., ZOLA, H., SHANNON, M. F. & BARRY, S. C. 2010. Genome-wide identification of human FOXP3 target genes in natural regulatory T cells. *J Immunol*, 185, 1071-81.
- SAFINIA, N., SCOTTA, C., VAIKUNTHANATHAN, T., LECHLER, R. I. & LOMBARDI, G. 2015. Regulatory T Cells: Serious Contenders in the Promise for Immunological Tolerance in Transplantation. *Front Immunol*, 6, 438.
- SAGE, P. T. & SHARPE, A. H. 2015. T follicular regulatory cells in the regulation of B cell responses. *Trends Immunol*, 36, 410-8.

- SAKAGUCHI, S. 2000. Regulatory T cells: key controllers of immunologic self-tolerance. *Cell*, 101, 455-8.
- SAKAGUCHI, S. 2003. The origin of FOXP3-expressing CD4(+) regulatory T cells: thymus or periphery. *Journal of Clinical Investigation*, 112, 1310-1312.
- SAKAGUCHI, S., YAMAGUCHI, T., NOMURA, T. & ONO, M. 2008. Regulatory T cells and immune tolerance. *Cell*, 133, 775-87.
- SALLUSTO, F., CASSOTTA, A., HOCES, D., FOGLIERINI, M. & LANZAVECCHIA, A. 2018. Do Memory CD4 T Cells Keep Their Cell-Type Programming: Plasticity versus Fate Commitment? T-Cell Heterogeneity, Plasticity, and Selection in Humans. *Cold Spring Harbor perspectives in biology*, 10, a029421.
- SALLUSTO, F., GEGINAT, J. & LANZAVECCHIA, A. 2004. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu Rev Immunol*, 22, 745-63.
- SAMBORSKI, P. & GRZYMISŁAWSKI, M. 2015. The Role of HSP70 Heat Shock Proteins in the Pathogenesis and Treatment of Inflammatory Bowel Diseases. *Adv Clin Exp Med*, 24, 525-30.
- SAMUSIK, N., GOOD, Z., SPITZER, M. H., DAVIS, K. L. & NOLAN, G. P. 2016. Automated mapping of phenotype space with single-cell data. *Nat Methods*, 13, 493-6.
- SÁNCHEZ-TILLÓ, E., SILES, L., DE BARRIOS, O., CUATRECASAS, M., VAQUERO, E. C., CASTELLS, A. & POSTIGO, A. 2011. Expanding roles of ZEB factors in tumorigenesis and tumor progression. *Am J Cancer Res*, 1, 897-912.
- SAUNDERS, C. J., ZHAO, W. & ARDINGER, H. H. 2009. Comprehensive ZEB2 gene analysis for Mowat-Wilson syndrome in a North American cohort: a suggested approach to molecular diagnostics. *Am J Med Genet A*, 149a, 2527-31.
- SCHAERLI, P., WILLIMANN, K., LANG, A. B., LIPP, M., LOETSCHER, P. & MOSER, B. 2000. CXC chemokine receptor 5 expression defines follicular homing T cells with B cell helper function. *The Journal of experimental medicine*, 192, 1553-1562.
- SCHALLER, T. H., BATICH, K. A., SURYADEVARA, C. M., DESAI, R. & SAMPSON, J. H. 2017. Chemokines as adjuvants for immunotherapy: implications for immune activation with CCL3. *Expert review of clinical immunology*, 13, 1049-1060.

- SCHLITZER, A., SIVAKAMASUNDARI, V., CHEN, J., SUMATOH, H. R., SCHREUDER, J., LUM, J., MALLERET, B., ZHANG, S., LARBI, A., ZOLEZZI, F., RENIA, L., POIDINGER, M., NAIK, S., NEWELL, E. W., ROBSON, P. & GINHOUX, F. 2015. Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nat Immunol*, 16, 718-28.
- SCHMIEDEL, B. J., SINGH, D., MADRIGAL, A., VALDOVINO-GONZALEZ, A. G., WHITE, B. M., ZAPARDIEL-GONZALO, J., HA, B., ALTAY, G., GREENBAUM, J. A., MCVICKER, G., SEUMOIS, G., RAO, A., KRONENBERG, M., PETERS, B. & VIJAYANAND, P. 2018. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*, 175, 1701-1715.e16.
- SCHUMANN, K., LIN, S., BOYER, E., SIMEONOV, D. R., SUBRAMANIAM, M., GATE, R. E., HALIBURTON, G. E., YE, C. J., BLUESTONE, J. A., DOUDNA, J. A. & MARSON, A. 2015. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences*, 112, 10437.
- SCOTT, C. L., SOEN, B., MARTENS, L., SKRYPEK, N., SAELENS, W., TAMINAU, J., BLANCKE, G., VAN ISTERDAEL, G., HUYLEBROECK, D., HAIGH, J., SAEYS, Y., GUILLIAMS, M., LAMBRECHT, B. N. & BERX, G. 2016. The transcription factor Zeb2 regulates development of conventional and plasmacytoid DCs by repressing Id2. *J Exp Med*, 213, 897-911.
- SCOTT, C. L., T'JONCK, W., MARTENS, L., TODOROV, H., SICHEN, D., SOEN, B., BONNARDEL, J., DE PRIJCK, S., VANDAMME, N., CANNOODT, R., SAELENS, W., VANNESTE, B., TOUSSAINT, W., DE BLESER, P., TAKAHASHI, N., VANDENABEELE, P., HENRI, S., PRIDANS, C., HUME, D. A., LAMBRECHT, B. N., DE BAETSELIER, P., MILLING, S. W. F., VAN GINDERACHTER, J. A., MALISSEN, B., BERX, G., BESCHIN, A., SAEYS, Y. & GUILLIAMS, M. 2018. The Transcription Factor ZEB2 Is Required to Maintain the Tissue-Specific Identities of Macrophages. *Immunity*, 49, 312-325.e5.
- SEKI, A. & RUTZ, S. 2018. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *Journal of Experimental Medicine*, 215, 985-997.
- SEKIDO, R., MURAI, K., FUNAHASHI, J., KAMACHI, Y., FUJISAWA-SEHARA, A., NABESHIMA, Y. & KONDOH, H. 1994. The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol Cell Biol*, 14, 5692-700.
- SERROUKH, Y., GU-TRANTIEN, C., HOOSHIAR KASHANI, B., DEFRANCE, M., VU MANH, T.-P., AZOUZ, A., DETAVERNIER, A., HOYOIS, A., DAS, J., BIZET, M., POLLET, E., TABBUSO, T., CALONNE, E., VAN GISBERGEN, K., DALOD, M.,

- FUKS, F., GORIELY, S. & MARCHANT, A. 2018. The transcription factors Runx3 and ThPOK cross-regulate acquisition of cytotoxic function by human Th1 lymphocytes. *eLife*, 7, e30496.
- SHEVACH, E. M. & THORNTON, A. M. 2014. tTregs, pTregs, and iTregs: Similarities and Differences. *Immunological reviews*, 259, 88-102.
- SHEVYREV, D. & TERESHCHENKO, V. 2020. Treg Heterogeneity, Function, and Homeostasis. *Frontiers in Immunology*, 10.
- SIMON, P. 2003. Q-Gene: processing quantitative real-time RT-PCR data. *Bioinformatics*, 19, 1439-40.
- SINGH, K., HJORT, M., THORVALDSON, L. & SANDLER, S. 2015. Concomitant analysis of Helios and Neuropilin-1 as a marker to detect thymic derived regulatory T cells in naïve mice. *Sci Rep*, 5, 7767.
- SLADE, C. D., REAGIN, K. L., LAKSHMANAN, H. G., KLONOWSKI, K. D. & WATFORD, W. T. 2020. Placenta-specific 8 limits IFN $\gamma$  production by CD4 T cells in vitro and promotes establishment of influenza-specific CD8 T cells in vivo. *PLoS one*, 15, e0235706-e0235706.
- STEPHENS, R. & LANGHORNE, J. 2010. Effector memory Th1 CD4 T cells are maintained in a mouse model of chronic malaria. *PLoS pathogens*, 6, e1001208-e1001208.
- STRAUSS, L., BERGMANN, C. & WHITESIDE, T. L. 2009. Human circulating CD4<sup>+</sup>CD25<sup>high</sup>Foxp3<sup>+</sup> regulatory T cells kill autologous CD8<sup>+</sup> but not CD4<sup>+</sup> responder cells by Fas-mediated apoptosis. *J Immunol*, 182, 1469-80.
- SUN, X., SOMADA, S., SHIBATA, K., MUTA, H., YAMADA, H., YOSHIHARA, H., HONDA, K., NAKAMURA, K., TAKAYANAGI, R., TANI, K., PODACK, E. R. & YOSHIKAI, Y. 2008. A critical role of CD30 ligand/CD30 in controlling inflammatory bowel diseases in mice. *Gastroenterology*, 134, 447-58.
- SUN, Y., LI, L., XIE, R., WANG, B., JIANG, K. & CAO, H. 2019. Stress Triggers Flare of Inflammatory Bowel Disease in Children and Adults. *Frontiers in Pediatrics*, 7.
- TAKAHASHI, T., TAGAMI, T., YAMAZAKI, S., UEDE, T., SHIMIZU, J., SAKAGUCHI, N., MAK, T. W. & SAKAGUCHI, S. 2000. Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J Exp Med*, 192, 303-10.



- TAN, T. G., MATHIS, D. & BENOIST, C. 2016. Singular role for T-BET+CXCR3+ regulatory T cells in protection from autoimmune diabetes. *Proc Natl Acad Sci U S A*, 113, 14103-14108.
- TANAKA, K., NAMBA, T., ARAI, Y., FUJIMOTO, M., ADACHI, H., SOBUE, G., TAKEUCHI, K., NAKAI, A. & MIZUSHIMA, T. 2007. Genetic evidence for a protective role for heat shock factor 1 and heat shock protein 70 against colitis. *J Biol Chem*, 282, 23240-52.
- THORNTON, A. M., KORTY, P. E., TRAN, D. Q., WOHLFERT, E. A., MURRAY, P. E., BELKAID, Y. & SHEVACH, E. M. 2010. Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory cells. *J Immunol*, 184, 3433-41.
- TIAN, Y., BABOR, M., LANE, J., SCHULTEN, V., PATIL, V. S., SEUMOIS, G., ROSALES, S. L., FU, Z., PICARDA, G., BUREL, J., ZAPARDIEL-GONZALO, J., TENNEKON, R. N., DE SILVA, A. D., PREMAWANSA, S., PREMAWANSA, G., WIJEWICKRAMA, A., GREENBAUM, J. A., VIJAYANAND, P., WEISKOPF, D., SETTE, A. & PETERS, B. 2017. Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nature communications*, 8, 1473-1473.
- TREBAK, M. & KINET, J.-P. 2019. Calcium signalling in T cells. *Nature reviews Immunology*, 19, 154-169.
- TRIFARI, S., KAPLAN, C. D., TRAN, E. H., CRELLIN, N. K. & SPITS, H. 2009. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from T(H)-17, T(H)1 and T(H)2 cells. *Nat Immunol*, 10, 864-71.
- UNIKEN VENEMA, W. T., VOSKUIL, M. D., VILA, A. V., VAN DER VRIES, G., JANSEN, B. H., JABRI, B., FABER, K. N., DIJKSTRA, G., XAVIER, R. J., WIJMENGA, C., GRAHAM, D. B., WEERSMA, R. K. & FESTEN, E. A. 2019. Single-Cell RNA Sequencing of Blood and Ileal T Cells From Patients With Crohn's Disease Reveals Tissue-Specific Characteristics and Drug Targets. *Gastroenterology*, 156, 812-815.e22.
- VAHEDI, G., KANNO, Y., FURUMOTO, Y., JIANG, K., PARKER, S. C. J., ERDOS, M. R., DAVIS, S. R., ROYCHOUDHURI, R., RESTIFO, N. P., GADINA, M., TANG, Z., RUAN, Y., COLLINS, F. S., SARTORELLI, V. & O'SHEA, J. J. 2015. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*, 520, 558-562.
- VAN GASSEN, S., CALLEBAUT, B., VAN HELDEN, M. J., LAMBRECHT, B. N., DEMEESTER, P., DHAENE, T. & SAEYS, Y. 2015. FlowSOM: Using self-organizing

- maps for visualization and interpretation of cytometry data. *Cytometry Part A*, 87, 636-645.
- VAN GRUNSVEN, L. A., Taelman, V., Michiels, C., Verstappen, G., Souopgui, J., Nichane, M., Moens, E., Opdecamp, K., Vanhomwegen, J., Kricha, S., Huylebroeck, D. & Bellefroid, E. J. 2007. X<sub>Sip1</sub> neutralizing activity involves the co-repressor CtBP and occurs through BMP dependent and independent mechanisms. *Dev Biol*, 306, 34-49.
- VAN HELDEN, M. J., Goossens, S., Daussy, C., Mathieu, A. L., Faure, F., Marcais, A., Vandamme, N., Farla, N., Mayol, K., Viel, S., Degouve, S., Debiens, E., Seuntjens, E., Conidi, A., Chaix, J., Mangeot, P., De Bernard, S., Buffat, L., Haigh, J. J., Huylebroeck, D., Lambrecht, B. N., Berx, G. & Walzer, T. 2015. Terminal NK cell maturation is controlled by concerted actions of T-bet and Zeb2 and is essential for melanoma rejection. *J Exp Med*, 212, 2015-25.
- VAN LEEUWE, T. M., Arentshorst, M., Ernst, T., Alazi, E., Punt, P. J. & Ram, A. F. J. 2019. Efficient marker free CRISPR/Cas9 genome editing for functional analysis of gene families in filamentous fungi. *Fungal Biology and Biotechnology*, 6, 13.
- VAN OOSTERWIJK, M. F., Juwana, H., Arens, R., Tesselaaar, K., Van Oers, M. H., Elderling, E. & Van Lier, R. A. 2007. CD27-CD70 interactions sensitise naive CD4<sup>+</sup> T cells for IL-12-induced Th1 cell development. *Int Immunol*, 19, 713-8.
- VANDEWALLE, C., Van Roy, F. & Berx, G. 2009. The role of the ZEB family of transcription factors in development and disease. *Cell Mol Life Sci*, 66, 773-87.
- VELDHOEN, M., Uyttenhove, C., Van Snick, J., Helmbjy, H., Westendorf, A., Buer, J., Martin, B., Wilhelm, C. & Stockinger, B. 2008. Transforming growth factor-beta 'reprograms' the differentiation of T helper 2 cells and promotes an interleukin 9-producing subset. *Nat Immunol*, 9, 1341-6.
- VERSCHUEREN, K., Remacle, J. E., Collart, C., Kraft, H., Baker, B. S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M. T., Bodmer, R., Smith, J. C. & Huylebroeck, D. 1999. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem*, 274, 20489-98.
- VERSTAPPEN, G., Van Grunsvan, L. A., Michiels, C., Van De Putte, T., Souopgui, J., Van Damme, J., Bellefroid, E., Vandekerckhove, J. & Huylebroeck, D. 2008. Atypical Mowat-Wilson patient confirms the importance

- of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. *Hum Mol Genet*, 17, 1175-83.
- VOLPE, E., SAMBUCCI, M., BATTISTINI, L. & BORSELLINO, G. 2016. Fas–Fas Ligand: Checkpoint of T Cell Functions in Multiple Sclerosis. *Frontiers in Immunology*, 7.
- WANG, F., WAN, L., ZHANG, C., ZHENG, X., LI, J. & CHEN, Z. K. 2009. Tim-3-Galectin-9 pathway involves the suppression induced by CD4+CD25+ regulatory T cells. *Immunobiology*, 214, 342-9.
- WANG, T., LIANG, Z. A., SANDFORD, A. J., XIONG, X. Y., YANG, Y. Y., JI, Y. L. & HE, J. Q. 2012. Selection of suitable housekeeping genes for real-time quantitative PCR in CD4(+) lymphocytes from asthmatics with or without depression. *PLoS One*, 7, e48367.
- WANG, Y., MISUMI, I., GU, A.-D., CURTIS, T. A., SU, L., WHITMIRE, J. K. & WAN, Y. Y. 2013. GATA-3 controls the maintenance and proliferation of T cells downstream of TCR and cytokine signaling. *Nature immunology*, 14, 714-722.
- WANG, Y. & ZHOU, B. P. 2013. Epithelial-mesenchymal Transition---A Hallmark of Breast Cancer Metastasis. *Cancer Hallm*, 1, 38-49.
- WEIGELIN, B., DEN BOER, A. T., WAGENA, E., BROEN, K., DOLSTRA, H., DE BOER, R. J., FIGDOR, C. G., TEXTOR, J. & FRIEDL, P. 2021. Cytotoxic T cells are able to efficiently eliminate cancer cells by additive cytotoxicity. *Nat Commun*, 12, 5217.
- WILLEMSSEN, R., RONTELTAP, C., HEUVELING, M., DEBETS, R. & BOLHUIS, R. 2005. Redirecting human CD4+ T lymphocytes to the MHC class I-restricted melanoma antigen MAGE-A1 by TCR alpha gene transfer requires CD8alpha. *Gene Ther*, 12, 140-6.
- WING, J. B. & SAKAGUCHI, S. 2012. Multiple treg suppressive modules and their adaptability. *Frontiers in Immunology*, 3, 178.
- WU, X., BRISENO, C. G., GRAJALES-REYES, G. E., HALDAR, M., IWATA, A., KRETZER, N. M., KC, W., TUSSIWAND, R., HIGASHI, Y., MURPHY, T. L. & MURPHY, K. M. 2016. Transcription factor Zeb2 regulates commitment to plasmacytoid dendritic cell and monocyte fate. *Proc Natl Acad Sci U S A*, 113, 14775-14780.
- XIA, Z. W., ZHONG, W. W., MEYROWITZ, J. S. & ZHANG, Z. L. 2008. The role of heme oxygenase-1 in T cell-mediated immunity: the all encompassing enzyme. *Curr Pharm Des*, 14, 454-64.

- XIE, M. M. & DENT, A. L. 2018. Unexpected Help: Follicular Regulatory T Cells in the Germinal Center. *Frontiers in Immunology*, 9.
- XU, L. X., GRIMALDO, S., QI, J. W., YANG, G. L., QIN, T. T., XIAO, H. Y., XIANG, R., XIAO, Z., LI, L. Y. & ZHANG, Z. S. 2014. Death receptor 3 mediates TNFSF15- and TNF $\alpha$ -induced endothelial cell apoptosis. *Int J Biochem Cell Biol*, 55, 109-18.
- YAMANO, T., STEINERT, M. & KLEIN, L. 2015. Thymic B Cells and Central T Cell Tolerance. *Frontiers in Immunology*, 6.
- YAMASHITA, Y., YATABE, Y., TSUZUKI, T., NAKAYAMA, A., HASEGAWA, Y., KOJIMA, H., NAGASAWA, T. & MORI, N. 1998. Perforin and granzyme expression in cytotoxic T-cell lymphomas. *Mod Pathol*, 11, 313-23.
- YANG, X. O., NURIEVA, R., MARTINEZ, G. J., KANG, H. S., CHUNG, Y., PAPPU, B. P., SHAH, B., CHANG, S. H., SCHLUNS, K. S., WATOWICH, S. S., FENG, X. H., JETTEN, A. M. & DONG, C. 2008. Molecular antagonism and plasticity of regulatory and inflammatory T cell programs. *Immunity*, 29, 44-56.
- YIN, S. Y., PENG, A. P., HUANG, L. T., WANG, Y. T., LAN, C. W. & YANG, N. S. 2013. The Phytochemical Shikonin Stimulates Epithelial-Mesenchymal Transition (EMT) in Skin Wound Healing. *Evid Based Complement Alternat Med*, 2013, 262796.
- ZANOLI, L., RASTELLI, S., INSERRA, G. & CASTELLINO, P. 2015. Arterial structure and function in inflammatory bowel disease. *World J Gastroenterol*, 21, 11304-11.
- ZHENG, S. G., WANG, J., WANG, P., GRAY, J. D. & HORWITZ, D. A. 2007. IL-2 is essential for TGF-beta to convert naive CD4<sup>+</sup>CD25<sup>-</sup> cells to CD25<sup>+</sup>Foxp3<sup>+</sup> regulatory T cells and for expansion of these cells. *J Immunol*, 178, 2018-27.
- ZHU, J. & PAUL, W. E. 2010. Peripheral CD4<sup>+</sup> T-cell differentiation regulated by networks of cytokines and transcription factors. *Immunol Rev*, 238, 247-62.