# A Portrait of $V\gamma 9V\delta 2$ *T cell* Exhaustion: Chronic Stimulation Models, RNA Transcriptomics and CRISPR Screens

**Meshael Mansour Alturki**
UCL

A thesis submitted to University College London (UCL) for the degree of
DOCTOR OF PHILOSOPHY

I, Meshael Alturki confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## Abstract

T cell exhaustion has been identified as a major challenge for T cell therapeutics. In conventional T cells (αβ T cells), the study of exhaustion has led to insights into its aetiology, and potential therapeutic approaches to overcome it. Our research focuses on a novel area of study-T cell exhaustion in human γδ T cells, particularly the Vγ9Vδ2 subtype, which is the most prevalent in peripheral blood. These cells hold therapeutic promise due to their broad cytotoxic characteristics and ability to bridge innate and adaptive immunity. However, the understanding of how exhaustion can affect the behaviour and functionality of these cells is yet to be established.

Our study involved *in vitro* chronic stimulation protocols using combinations of zoledronic acid (ZOL) and OKT3 (anti-CD3 monoclonal antibody) or Daudi target cells, both stimulating via the Vγ9Vδ2 T cell receptor. The findings from our ZOL model, which showed a notable decrease in absolute T-cell numbers and significant upregulation of genes associated with T cell exhaustion, could have significant implications for understanding and potentially overcoming T cell exhaustion in Vγ9Vδ2 T cells.

Bulk-RNA sequencing of Vγ9Vδ2 T cells identified distinct signatures for the multiple stimulated samples, and a list of highly differentiated genes was selected for further investigation by a pooled CRISPR screen. We optimised and tested a suitable protocol for a CRISPR screen in Vδ2 T cells. Two gRNA hits were enriched in the ZOL and multiple stimulated samples (targets for SLC37A3 and GORASP2 genes).

In conclusion, a simple *in vitro* assay was developed to study hypofunctionaity in Vδ2 T cells, allowing studies to evaluate RNA profiles and regulatory mechanisms of exhaustion induction in this cell type. Finally, selected genes were screened in a CRISPR-based pooled screen and hit gRNAs from stimulated Vδ2 T cells were identified.

## Impact Statement

The work presented in this thesis is novel to the field of gamma delta T cells as it investigates the effects of chronic stimulation models in inducing exhaustion in Vδ2 T cells. As this has not been investigated before, different models of chronic stimulation have been tested, and functionality assays have also been evaluated to aid in understanding the impact of chronic stimulation on Vδ2 T cells. Moreover, RNA profiling to differentially stimulated Vδ2 T cell conditions was performed to ensure proper phenotyping of exhaustion and T cell activation markers at the transcriptomic levels. Finally, an optimised Lentiviral CRISPR screen protocol was developed to test the knock-out effects of multiple hits from the RNA profiling. The studies concluded that two potential gRNA hits were enriched with multiple stimulations targeting the T cell receptor, these were SLC37A3 and GORASP2. This thesis has identified a model to stimulate Vδ2 T cells chronically, showed RNA profiles of Vδ2 cells that were chronically stimulated in addition to developing a protocol for CRISPR screens in Vδ2 T cells and performing CRISPR screens with the identification of two potential genetic mediators of hypo-proliferation in the context of chronic stimulation. Identified hits will require further assessments as single knockouts in Vδ2 T cells to conclude their specific roles in Vδ2 T cell exhaustion.

## Acknowledgements

I would like to express my deepest gratitude to all those who supported me throughout my PhD journey.

First and foremost, I must thank my supervisor, John Anderson, for his guidance, advice, and unwavering support at every step of my project. His encouragements and belief in my abilities were crucial to my success.

Secondly, I extend my heartfelt thanks to every member of the Anderson group. You have been an integral part of my PhD experience, contributing to some of my fondest memories.

To my friends and family, your unwavering support and encouragements sustained me through the challenges of my PhD journey and the stressful times brought about by the COVID-19 pandemic. Thank you for standing by me and never wavering in your belief in me.

I am also profoundly thankful to my country for providing me with the opportunity to pursue my studies and for its ongoing support.

Lastly, I am immensely grateful to everyone who has been a part of my journey. This achievement would not have been possible without you.

# Table of Contents

**LIST OF ABBREVIATIONS** ..... **8**

**CHAPTER 1** ..... **14**

INTRODUCTION ..... 14
*Exhaustion* ..... *14*
*Acute vs Chronic infection* ..... *14*
*Advances in exhaustion studies* ..... *15*
*The landscape of exhaustion* ..... *16*
*Key transcriptional and epigenetic markers* ..... *17*
*Reversal of exhaustion* ..... *22*
*Immunotherapy platforms* ..... *23*
*Gamma delta (γδ) T cells* ..... *23*
*Unique characteristics of γδ T cells* ..... *24*
*Synthetic Biology approaches to study exhaustion in γδ T cells.* ..... *26*
*Precision and accuracy in CRISPR technologies* ..... *32*
*CRISPR off-target effects* ..... *33*
*CRISPR Delivery* ..... *33*
*Aims and Objectives* ..... *34*

**CHAPTER 2** ..... **36**

MATERIALS AND METHODS ..... 36
*Isolation of primary human peripheral blood mononuclear cells* ..... *36*
*T cells activation and phenotyping* ..... *36*
*Staining* ..... *37*
*E changeProliferation assay* ..... *38*
*Cell death and apoptosis assay* ..... *39*
*Fluorescence-activated cell sorter* ..... *39*
*List of Panels* ..... *42*
*Functional Assays* ..... *46*
*Transcriptomic analysis of Vγ9δ2 T cells* ..... *49*
*CRISPR Knock-out Screen on Vδ2 T cells* ..... *51*
*Cell Gating Strategy* ..... *57*

**CHAPTER 3** ..... **59**

CHRONIC STIMULATION MODELS TO STUDY EXHAUSTION IN Vδ2 T CELLS ..... 59
*Introduction and Summary* ..... *59*
*Developed in vitro Stimulation Models* ..... *60*
*OKT3 based chronic Co-stimulation of Vγ9Vδ2 and αβ T cells* ..... *63*
*Zoledronic acid and OKT3-based chronic stimulation of Vγ9Vδ2 T cells* ..... *70*
*Effect of Daudi chronic stimulation model on cell numbers and TCR internalization* ..... *79*
*Conclusion* ..... *86*

**CHAPTER 4** ..... **91**

FUNCTIONAL STUDIES USING SINGLE AND DOUBLE-STIMULATED Vδ2 T SHOW OVERLAPPING RESPONSES TO VARIOUS CANCER TARGETS. ..... 91
*Introduction* ..... *91*
*Co-culture using Daudi and Vδ2 T cells* ..... *92*
*CFSE tracking in Vδ2 T cells* ..... *101*
*Re-stimulation assay of Vδ2 T cells with target cells* ..... *104*
*Conclusion* ..... *108*

**CHAPTER 5** ..... **110**

RNA TRANSCRIPTOMICS OF Vγ9Vδ2 T CELLS ..... 110
*Introduction* ..... *110*
*Experiment 1: RNA profiling of Vδ2 T cells* ..... *111*
*Gene set enrichment analysis: Enrichment maps* ..... *125*
*Experiment 2: RNA profiling of Vδ2 T cells* ..... *132*
*Shared genes between the two transcriptomics data sets* ..... *154*
*Conclusion* ..... *156*

**CHAPTER 6** ..... **158**

CRISPR SCREENS ON Vδ2 T CELLS ..... 158
*Introduction* ..... *158*
*CRISPR Knockout Screen on Vδ2 T cells* ..... *160*
*Use of Different Transfection Agents to Increase Viral Titres* ..... *164*
*Pilot Expansion for Increased GFP Enrichment Following Transduction* ..... *169*
*Expansion with optimised time-line* ..... *172*
*Activation of target genes using CRISPR dCasmini system* ..... *184*
*Conclusion* ..... *187*

**CHAPTER 7** ..... **189**

DISCUSSION AND CONCLUSIONS ..... 189
FUTURE DIRECTIONS ..... 199
POSITIVE OUTCOMES OF THE STUDY ..... 200

**REFERENCES** ..... **201**

**APPENDIX** ..... **214**

FULL R SCRIPT FOR RNA PROFILING ANALYSIS ..... 214
GRNA LIBRARY FOR CRISPR KNOCK-OUT SCREEN ..... 234

# List of Abbreviations

| Abbreviation | Definition |
|---|---|
| αβ | Alpha-beta |
| γδ | Gamma-delta |
| 293T | Human embryonic kidney cell line HEK293 |
| AAV | Adeno-associated virus |
| AdV | Adenoviruse |
| AP-1 | Activator Protein 1 |
| B30.2 | Intracellular signalling domain of BTN3A1 |
| BATF | Basic Leucine Zipper ATF-Like Transcription Factor |
| BATF3 | Basic Leucine Zipper ATF-Like Transcription Factor 3 |
| BCL-6 | B-Cell Lymphoma 6 Protein |
| BLIMP-1 | B Lymphocyte-Induced Maturation Protein 1 |
| BTN2A1 | Butyrophilin Subfamily 2 Member A1 |
| BTN3A1 | Butyrophilin Subfamily 3 Member A1 |
| bZIP | Basic Leucine Zipper Domain |
| C57BL/6 | C57 black 6 strain |
| CAR | Chimeric antigen receptor |
| Cas | CRISPR-associated protein |
| Cas12a | CRISPR-associated protein derived from the Lachnospiraceae bacterium |
| Cas12f | CRISPR-associated protein derived from uncultivated archaea |
| CCL22 | C-C motif chemokine 22 |
| CD28 | T-cell-specific surface glycoprotein CD28 |
| CD300A | Cluster of differentiation 300 A |
| CD40L | Cluster of Differentiation 40 ligand (type II transmembrane glycoprotein) |
| CD40LG | Gene expressing CD40L |
| CD69 | Cluster of Differentiation 69 |
| CD80 | Cluster of Differentiation 80 |
| CDR3 | Complementarity determining region 3 |
| CHAC1 | Glutathione specific gamma-glutamylcyclotransferase 1 |
| cJUN | Jun Proto-Oncogene |
| Cl-13 | Clone-13 Armstrong viral strain |
| Cnet | Concept network |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CRISPRa | CRISPR activation |
| CRISPRi | CRISPR interference |
| CSF2 | Colony-stimulating factor 2 |

| | |
|---|---|
| CTL | Cytotoxic T-lymphocytes |
| CTLA-4 | Cytotoxic T-lymphocyte associated protein 4 |
| CTV | CellTrace™ Violet |
| CXC3CR1 | CX3C motif chemokine receptor 1 |
| CXCR5 | C-X-C chemokine receptor type 5 |
| dCas | Dead or deactivared Cas |
| DMSO | Dimethylsulfoxide |
| DNA | Deoxyribonucleic acid |
| DSB | Double-strand break |
| dsDNA | Double stranded Deoxyribonucleic acid |
| EBF4 | Transcription factor for the Olf/EBF family of helix-loop-helix transcription factors |
| EF-1 α | Human elongation factor-1 alpha |
| Env | Viral gene encoding envelope protein |
| EOMES | Eomesodermin |
| F1-ATPase | F-Type ATPase |
| FACS | Fluorescence-activated cell sorting |
| FCS | Fetal calf serum |
| FDA | Food and Drug Administration |
| FDR | False discovery rate |
| FOS | Proto-oncogene |
| FPPS | Farnesyl diphosphate synthase |
| Gag | Group-specific antigen and polyprotein with core structural proteins of the virus |
| GFP | Green fluorescent protein |
| GM-CSF | Immune modulatory molecule |
| GO | Gene Ontology |
| GORASP 2 | Golgi reassembly-stacking protein 2 |
| gRNA | Guide Ribonucleic acid |
| GSEA | Gene set enrichment analysis |
| HAVCR2 | T-cell immunoglobulin and mucin-domain containing-3 also known as TIM- |
| HDR | Homology directed repair |
| HMG protein | High-Mobility Group protein |
| HMOX1 | Heme oxygenase 1 gene |
| hMSH2 | mismatch repair genes |
| ID3 | Inhibitor Of DNA Binding 3 |
| IFIT1 | Interferon induced protein with tetratricopeptide repeats 1 |
| IFIT2 | Interferon induced protein with tetratricopeptide repeats 2 |
| IFIT3 | Interferon induced protein with tetratricopeptide repeats 3 |
| IFN-γ | Interferon-gamma |
| IFN-α | Interferon-alpha |

| | |
|---|---|
| IL-10 | Interleukin 10 |
| IL-12 | Interleukin 12 |
| IL-17 | Interleukin 17 |
| IL-2 | Interleukin 2 |
| IL-4 | Interleukin 4 |
| IL17F | Interleukin 17F |
| IPP | Isopentenyl diphosphate |
| IRF4 | Interferon Regulatory Factor 4 |
| LAG-3 | Lymphocyte-activation gene 3 |
| LB | Lysogeny broth |
| LEF1 | Lymphoid Enhancer Binding Factor 1 |
| LentiX-293T | Human embryonic kidney cell line (HEK 293) cell line for enhanced viral protein production |
| LPS | Lipopolysaccharide |
| Ly108 | Signaling lymphocyte activation molecule family 6 (slamf6) |
| MAPK11 | Mitogen-activated protein kinase 11 |
| MFI | Mean fluorescence intensity |
| MHC | Major Histocompatibility complex system |
| HLA | Human leukocyte antigen |
| mRNA | Messenger Ribonucleic acid |
| msigDB | Molecular Signatures Database |
| MXI1 | MAX interactor 1 |
| MYC | Proto-oncogene |
| NDRG1 | N-myc downstream-regulated gene |
| NES | Normalized enrichment score |
| NFAT | Nuclear factor of activated T cells |
| NGS | Next-generation sequencing |
| NHEJ | Non-homologous end joining |
| NK cells | Natural killer cells |
| NKG2D | Killer cell lectin-like receptor subfamily K, member 1 (Klrk1) |
| NR4A | Nuclear receptor subfamily 4 group a) |
| OKT3 | Anti-CD3 monoclonal, clone OKT3 |
| ORA | Over-representation analysis |
| p.f.u | Plaque-forming units |
| PAM | Protospacer adjacent motif |
| PARP9 | Poly (ADP-ribose) polymerase family member 9 |
| PBMCs | Peripheral blood mononuclear cells |
| PBS | Phosphate-buffered saline |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| PD-1 | Programmed cell death protein 1 |

| | |
|---|---|
| PD-L1 | Programmed death-ligand 1 |
| PDGFA | Platelet-derived growth factor receptor alpha |
| PEI | Polyethylenimine |
| PI | Propidium iodide |
| Pol | A gene in retroviruses entering target cells |
| PRKCE | Protein kinase C epsilon |
| QC | Quality control |
| R | Programming language |
| RARRES3 | Retinoic acid receptor responder protein |
| Rcf | Relative centrifugal force |
| RD114 | Feline endogenous gamma retrovirus |
| RDpro | Gamma retrovirus with a modified cytoplasmic tail |
| RhoB GTPase | Member of the Ras super-family and Rho sub-family of gtpases |
| RNA | Ribonucleic acid |
| RNP | Ribonucleoprotein |
| RSAD2 | Radical S-adenosyl methionine domain containing 2 |
| RXRA | Retinoid X receptor alpha |
| Seq | Sequencing |
| sgRNA | Single guide Ribonucleoprotein |
| slamf6 | Signaling lymphocyte activation molecule family 6 |
| SLC37A3 | Solute carrier family 37-member 3 |
| SLFN5 | Schlafen family member 5 |
| SpCas9 | Streptococcus pyogenes derived Cas9 |
| T-bet | T-box transcription factor TBX21 |
| TBX21 | T-box transcription factor T also known as T-bet |
| TCF-1 | T cell factor 1 |
| TCF12 | T cell factor 12 |
| TCF4 | T cell factor 4 |
| TCF7 | Gene encoding TCF-1 |
| TCF7L2 | Transcription factor 7-like 2 |
| TCR | T cell receptor |
| TEMRA | Terminally differentiated effector cells |
| Texprog1 | Exhausted progenitor 1 T cells |
| Texprog2 | Exhausted progenitor 2 T cells |
| TGFB1 | Transforming growth factor beta 1 |
| TGFBR3 | Transforming growth factor beta receptor III |
| TH1 | Type 1 T helper cells |
| TH2 | Type 2 T helper cells |
| TIGIT | T cell immunoreceptor with Ig and ITIM domains |

| | |
|---|---|
| TIM-3 | T-cell immunoglobulin and mucin-domain containing-3 |
| TMEM200B | Transmembrane protein 200B |
| TNF | Tumor necrosis factor |
| TOX | Thymocyte selection-associated high mobility group box protein |
| TOX1 | Thymocyte selection-associated high mobility group box protein 1 |
| TOX2 | Thymocyte selection-associated high mobility group box protein 2 |
| TOX3 | Thymocyte selection-associated high mobility group box protein 3 |
| TSST | Toxic shock syndrome toxin |
| Vγ9Vδ2 | Gamma 9 delta 2 T cell subtype |
| VPR | Transcription activation domain |
| VSV-g | Vesicular stomatitis virus G |
| WNT | Glycoprotein ligand that activates receptor-mediated signaling pathways |
| WNT10B | Wnt family member 10b |
| Zol | Zoledronic acid or Zoledronate |

# Chapter 1

## Introduction

### Exhaustion

The concept of T cell exhaustion has evolved rapidly in recent years and has been defined differently among different research fields. Initially, "exhaustion" was used to describe T cells in a setting of chronic infection. Moreover, many studies on T cell exhaustion were performed on a Lymphocytic choriomeningitis virus (LCMV) infection model in vivo. However, recent similar observations of T-cell exhaustion have focused on tumour studies (Blank et al., 2019). Even with the latest knowledge on exhaustion, there remains to be clarity regarding its meaning and definition.

### Acute vs Chronic infection

In the presence of foreign antigens, an army of cells actively kills and eradicates the disease. During a persistent infection, such as viral infections, T cells get over-activated and become hypofunctional, where they cannot control the disease, lately termed exhausted. In contrast, in normal immune responses and acute infections, the predominant subtypes of T cells are effector and memory T cells. From our understanding of the differentiation of alpha-beta (αβ) memory T cells, it is thought that naïve T cells are differentiated into memory cells and then into terminal effector cells. Importantly, all types of memory cells (i.e., Stem-like, central and effector memory) have effector functions. Terminally differentiated effector cells (i.e., TEMRA) lose their stem cell properties but are characterised by enhanced killing and cytokine production. TEMRA cells are thought to be short-lived. The balance between memory and effector cells is essential for the homeostasis of an immune response. Exhaustion is a homeostasis mechanism evolved to limit immunopathology during chronic infections when effector T cells are challenged with a persistent antigen load and reduce their effector functionality (Blank et al., 2019; McLane et al., 2015)

## Advances in exhaustion studies

Our understanding of exhaustion is in continuous development, especially with advancing research technologies and the use of high-resolution methods (i.e., sequencing on RNA and chromatin levels). T cells' exhaustion phenotype has been associated with many changes to their transcription and epigenetic program, distinguishing them from effector and memory cells (Khan et al. 2019; Wherry & Kurachi 2015). Exhausted T cells are of a distinct cell lineage committed to a state of hypofunctionality (Khan et al., 2019; Wherry & Kurachi, 2015). Moreover, a large population of exhausted T cells is a key indicator of chronic infection (Khan et al., 2019).

Crucial questions regarding the development of exhausted T cells are: why does it occur and what is its role in controlling infections? Suggestions imply that exhaustion might be a developmental process that has evolved to protect hosts from hyperactivation and immunopathology (Mann & Kaech, 2019).

An early comparative study on the impact of different LCMV-derived viral strains on C57 black 6 strain (C57BL/6 ) mice to induce persistent infections was performed by Moskophidis and co-workers (Moskophidis et al., 1995). They looked at the effects of different LCMV-derived viral strains on C57BL/6 mice. The aim was to measure the induction levels of persistent infections (Moskophidis et al., 1995). The strains included in the study are LCMV-Armstrong (wild-type), Clone-13 Armstrong (Cl-13) (derived from LCMV-Armstrong), and LCMV-Docile. Several findings regarding the correlation of each infection to the response of cytotoxic CD8+ T cells (CTL) were reported. First, CTLs easily maintained the clearance of the infection with different doses of LCMV-Armstrong. Second, with Cl-13 infection and using doses higher than or equal to $10^4$ plaque-forming units (p.f.u), a temporary persistence of the virus was observed and accompanied by a partial loss of the antiviral CTLs, causing a state of chronic infection. Finally, the LCMV-Docile strain was the most persistent at doses higher or equal to $10^4$ p.f.u. Docile infection was also associated with a long-lasting loss of antiviral CTLs. Using high viral doses resulted in the exhaustion and functional impairment of the antiviral CTLs.

LCMV-Armstrong and Cl-13 did not cause exhaustion in CTLs. Instead, they triggered lethal lymphocytic choriomeningitis and the development of immunopathology. Studies on the effects of different LCMV models on T cell phenotypes suggest that with more aggressive infections (i.e., Docile), the proportion of T cells entering a state of exhaustion is increased to prevent immune pathology and maintain a prolonged control of diseases. In summary, these early studies in chronic infection identified that exhausted T cells have evolved as a unique lineage of cells with specialised functions in controlling chronic infections and prevention of immunopathology (Moskophidis et al., 1995)

## The landscape of exhaustion

The functional and phenotypical landscape of exhausted T cells is yet to be fully understood. However, recent reports suggested two main classes of exhausted T cells: precursor-exhausted T cells and terminally exhausted T cells. Precursor-exhausted T cells, in common with memory precursor effector cells, highly express the transcription factor T cell factor 1 (TCF-1), and they also display stem cell characteristics such as the ability to self-renew and proliferate. Additionally, they have a role in generating terminally exhausted T cells (Mann & Kaech, 2019; Utzschneider et al., 2016). Moreover, clinical studies identified an association between TCF-$1^{+}$ programmed cell death protein $1^{+}$ (PD-$1^{+}$) memory-like CD8+ tumour infiltrating lymphocytes (TILs) and improved tumour control (Siddiqui et al., 2019). The use of checkpoint blockade, the inhibition of checkpoint proteins using antibodies to modulate T cell functionality, was incorporated into the study of T cell differentiation, in detail, it was shown that with checkpoint blockade of PD-1 and cytotoxic T-lymphocyte associated protein 4 (CTLA-4), less-differentiated memory-like CD8+ T cells produce a pool of effector CD8+ T cells with stem-like functions, causing the tumour control benefits (Siddiqui et al., 2019). Terminally exhausted T cells are identified with high levels of PD-1, T-cell immunoglobulin and mucin-domain containing-3 (TIM-3), and Thymocyte selection-associated high mobility group box protein (TOX). They are also recognised for their loss of functionality and inability to respond to checkpoint inhibitor treatments (Khan et al., 2019).

Additional factors such as T-box transcription factor TBX21 (T-bet) and Eomesodermin (EOMES) of the T-box transcription factors family were recently discovered as exhaustion-associated proteins. They have primary roles in the activation and differentiation of T cells and memory development (Wherry & Kurachi, 2015). T-bet and EOMES are expressed during the development of exhausted T cells and are essential for identifying and differentiating between the precursor-exhausted and terminally exhausted T cells phenotype. In short, a high T-bet to EOMES ratio has been proposed as a marker of progenitor T cells subset, while a low T-bet to EOMES ratio is associated with terminally exhausted T cells (Figure 1.1).

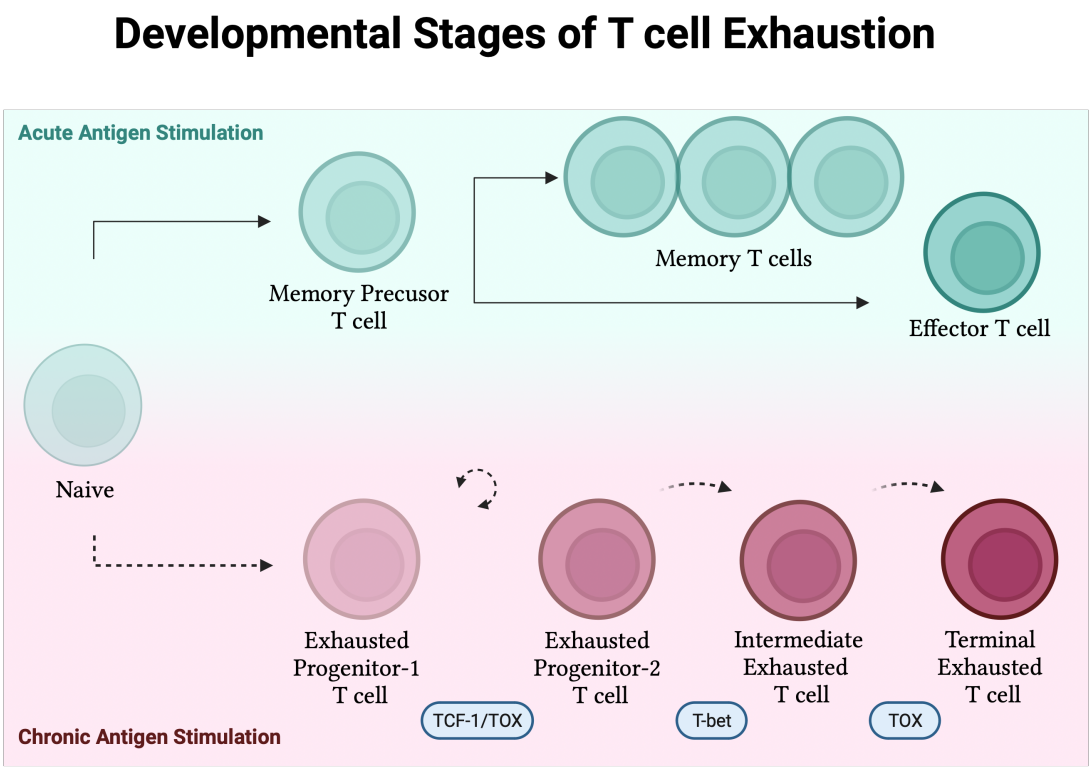



*Figure 1.1: Developmental stages of T cell exhaustion during acute or chronic antigen stimulation.*

## Key transcriptional and epigenetic markers

Recent advances in exhaustion research improved our definitions of T cell dysfunction and exhaustion. In addition, the ability to distinguish transcriptional and epigenetic signatures associated with exhaustion has enhanced our understanding of this phenomenon and insight into

how previously determined classical markers of exhaustion correlate with newly discovered transcription factors (Blank et al., 2019). One of the prominent transcription factors reported is TOX (Thymocyte selection-associated high mobility group box protein). TOX is known to function as the main moderator of T cell exhaustion. In a study by Khan and co-workers (Khan et al., 2019), a characterisation of T cell exhaustion in an LCMV model was accomplished. The report showed that TOX expression was very high in exhausted T cells. Moreover, TOX expression was sustained throughout the activation period (J. S. Chen & Eisenbarth, 2019; Sekine et al., 2020; Seo et al., 2019b). So far, the known characteristics of TOX are that it is expressed at an early stage during exhaustion and can trigger a commitment of T cells to the exhaustion pathway; it is also reported as an epigenetic regulator and modulates the chromatin accessibility of a network of exhaustion transcription factors (Khan et al., 2019) Even though it is shown that TOX levels remain high throughout an infection, its specific role at that stage is not precisely understood.

Further work by Beltra and co-workers (Beltra et al., 2020) provided a better overview of cellular dysfunction and the development of exhaustion in T cells by identifying four different developmental stages of exhausted T cells. The group used Ly108 (Signaling lymphocyte activation molecule family 6 (slamf6)) and CD69 (Cluster of Differentiation 69) to identify and characterise the different exhaustion stages of T cells in an LCMV infection model. The initial two stages can interchange and lead to the development of subsequent stages. Stage 1 is defined by high expression levels of TCF1, encoded by *TCF7*, and high TOX levels. These cells are termed exhausted progenitor 1 ($Tex^{prog1}$) Ly108+ CD69+ T cells. The second development phase starts with a decreased TCF-1 expression, and with maintained high TOX levels, the cells are named exhausted progenitor 2 ($Tex^{prog2}$) $Ly108^{+}$ $CD69^{-}$ T cells. The third developmental stage, where exhausted intermediate $Ly108^{-}$ CD69 T cells are developed, is initiated by the loss of TCF-1 expression, a decrease in TOX levels and the acquisition of high T-bet expression levels. The cells at the final stage of exhaustion development are identified as $Ly108^{-}$ $CD69^{+}$ terminally exhausted T cells. At this stage, the cells are associated with a negative expression of TCF-1, a low expression of T-bet, and a high TOX expression. Terminally exhausted cells in this scheme also display a gain of high expression levels of the transcription factor EOMES (Beltra et al., 2020)

NFAT (nuclear factor of activated T cells) is an essential transcription factor, especially in driving T-cell signalling by binding to transcription co-factors. Naturally, TCR (T cell receptor) stimulation induces the formation of NFAT and AP-1(activator protein 1) complexes, which drive IL-2 (interleukin-2) transcription. In detail, NFAT-AP-1 Complexes are formed by increased levels of NFAT production from TCR signalling and AP-1 induced by costimulation, resulting in transcriptional T-cell activation and IL-2 production. Alterations to the balance of transcription co-factors with NFAT lead to different phenotypes. Increased levels of partner-less NFAT or complexes of BATF

(Basic leucine zipper transcription factor) with NFAT are believed to drive exhaustion (Bengsch & Wherry, 2015).

A report by Man and co-workers (Man et al., 2017) recognised additional transcription factors contributing to exhaustion. Compared to functional T cells, increased expression of IRF4 (Interferon Regulatory Factor 4) and BATF, which are transcription factors known to be induced by TCR signalling, was observed in exhausted T cells. High levels of IRF4 and BATF increased the expression of PD-1 and other inhibitory receptors, impairing T cells' functionality. This paper also identified factors such as NFATc1 and NFATc2 (TCR-induced transcription factors) to be highly expressed in exhausted T cells and essential in inducing IRF4 levels. Hence, a coordinating role of IRF4 and NFAT proteins was discovered and explained by positive feedback demonstrating their function in establishing an exhaustion phenotype and suppression of memory T cells. The report also tested the effects of suppressing IRF4 and demonstrated that this manipulation reduced functional exhaustion and increased TCF-1 positive progenitor T cells thereby emphasising its contribution to generating functional exhaustion (Man et al., 2017).

There has been rapid progress toward the discovery of novel exhaustion-associated transcription factors. Some top targets recognised are TOX, BLIMP-1 (B Lymphocyte-Induced Maturation Protein 1), ID2 (Inhibitor Of Differentiation 2) and NR4A (Nuclear Receptor Subfamily 4 Group A), all of which have exhaustion enhancement functions (Bengsch & Wherry, 2015; Man et al.,

2017; Lynn et al., 2019). In contrast, there are several transcription factors with known association with precursor exhausted T cells, including TCF-1, cJUN (Jun Proto-Oncogene), LEF1 (Lymphoid Enhancer Binding Factor 1), BCL-6 (B-Cell Lymphoma 6 Protein), ID3 (Inhibitor Of DNA Binding 3) and SLAMF6 all of which have roles in exhaustion reversal and regeneration of effector functionality (Kallies et al., 2020; Shan et al., 2021; Wu et al., 2016).

NR4A proteins are critical regulators of T-cell exhaustion (J. Chen et al., 2019). They are present at very high levels in exhausted T cells, and their family of proteins include NR4A1, NR4A2 and NR4A3. ATAC-Sequencing (assay for transposase-accessible chromatin with sequencing) of CD8+ tumour infiltrating lymphocytes and CARs displayed enrichment of NR4A-binding motifs in chromatin accessibility regions, emphasising the role of NR4A genes in T cell exhaustion. The authors also showed the effects of knocking out all NR4A genes (NR4A1, NR4A2 and NR4A3) from T cells, and the results were an increased expression of NFAT and AP-1 proteins. Furthermore, investigating the relationship between NFAT and AP-1 protein expression in NR4A knockout cells, the authors found that the NFAT and AP-1 family of proteins, including Fos and Jun, are critical in maintaining T cells effector functionality (J. Chen et al., 2019). A positive feedback loop between NFAT, TOX and NR4A has been demonstrated and highlights the interplay between the three transcription factors to drive T cell exhaustion (J. Chen et al., 2019; Seo et al., 2019b).

Canonical AP-1 (c-JUN and c-FOS) family proteins are associated with pro-inflammation and T cell activation. In exhausted cell subtypes, Lynn and co-workers (Lynn et al., 2019) described a dramatic increase of BATF and BATF3 (Basic Leucine Zipper ATF-Like Transcription Factor 3), non-canonical AP-1 proteins with inhibitory effects, in exhausted T cells. High levels of the non-canonical AP-1-inhibitory proteins bZIP (Basic Leucine Zipper Domain) and IRF4 were observed to negatively control the Jun/FOS protein levels of the AP-1 family and consequently reduce T cell functionality. The authors showed significantly improved functionality and potency of CD4+ and CD8+ T cells that over-express c-JUN within multiple tumour models and were further validated by an increased expression of IL-2 and IFN-γ (Interferon-gamma). A range of exhaustion markers expressed in conventional T cells is shown in Table 1.1.

*Table 1.1: List of markers expressed during T cells and exhaustion development(Kallies et al., 2020).*

| Molecule | Function | Expression | |
|---|---|---|---|
| **Inhibitory receptors** | | **$T_{PEX}$** | **$T_{EX}$** |
| **TIM-3** | Inhibitory receptor | - | ++ |
| **PD-1** | Inhibits expansion of T cells | +/- | + |
| **LAG-3** | Negatively regulate cell cycle | ++ | ++ |
| **Surface Molecules** | | | |
| **Ly108 (slamf6)** | Unknown | ++ | - |
| **CXCR5** | Chemokine receptor | +/- | - |
| **Cytokines and cytotoxic molecules** | | | |
| **IL-2** | Stimulate proliferation and differentiation of T cells | +/- | - |
| **IFN-γ** | Effector cytokine | +/- | +/- |
| **TNF** | Effector cytokine | +/- | +/- |
| **Transcription factors** | | | |
| **TCF-1** | Promotion of $T_{PEX}$ generation | ++ | - |
| **ID3** | Promotes memory T cells | ++ | - |
| **TOX** | Regulate $T_{EX}$ at transcriptional level | +/- | + |
| **BCL-6** | Promotion of $T_{PEX}$ generation | + | - |
| **BLIMP-1** | Promote $T_{EX}$ and repress $T_{PEX}$ | - | ++ |
| **NR4A** | Promote $T_{EX}$ and repress $T_{PEX}$ | - | ++ |
| **T-bet** | Maintains $T_{EX}$ | + | + |
| **EOMES** | Maintains $T_{EX}$ | + | +/- |

## Reversal of exhaustion

There have been many efforts to restore the functionality of exhausted T cells during infections. CTLA-4 blockade was the first approved checkpoint inhibitor for cancer (Phan et al., 2003) Blocking PD-1 with antibodies has been very effective in restoring the functionality of T cells in cancer and has been approved for its clinical use in cancer therapy. Double blockade of CTLA-4 and PD-1 (Larkin et al., 2015) has demonstrated even greater clinical success. It showed great therapeutic potential, influencing the increased interest in evaluating new targets such as LAG-3 (Lymphocyte-activation gene 3) (Tawbi et al., 2022) and TIGIT (T cell immunoreceptor with Ig and ITIM domains) (Guillerey et al., 2018). Recent attempts at the rescue of exhausted T cells include the overexpression of T cells proliferation-promoting and activating genes such as C-Jun (Lynn et al., 2019) and TCF-1 (Shan et al., 2020) or the knockout of exhaustion-boosting transcription factors like TOX (Khan et al., 2019; Seo et al., 2019) and NR4A proteins (J. Chen et al., 2019). A double knock-out of BLIMP-1 and NR4A3 enhanced CAR T cell functionality by restoring *TCF7*-dependent T cell stemness (Jung et al., 2022).

Observations highlight the importance of TCF-1 in maintaining the precursor exhaustion phenotype of T cells (Wu et al., 2016). The group showed how the over-expression of TCF-1 reduces exhaustion in T cells and induces the generation of progenitor T cells. TCF-1 expression is also essential for the long-term maintenance of T cell effectivity, which further validates its role as a potential exhaustion-reversing target. Additionally, Wu, T. et al. found that TCF-1 overexpression preserves TIM-3-low progenitor T cells by increased levels of Bcl-6, stressing the cooperativity between TCF-1 and Bcl-6 and their benefits to the functionality of T cells. In contrast, Wu T. et al. also showed that the enrichment of type one IFN-γ genes in progenitor T cells suppresses the TCF-1-Bcl-6 pathway and T cell stemness at an early stage of chronic infection (Wu et al., 2016).

## Immunotherapy platforms

The field of adoptive T-cell therapy for cancer is one of the most promising in medical and clinical research. Thus far, it has produced major technologies that are revolutionary and currently used clinically. However, the predominant platform used for current studies and clinical trials is alpha-beta (αβ) T cells. Undoubtedly, there is a great amount of information and established experimentation relating to αβ T cells. However, there is a need to improve our current therapies by accelerating the discovery of new potential therapeutic candidates. In recent years, there has been an increased interest in examining various cell types for their possible use in immunotherapies. Examples include NK cells (natural killer cells) (Albinger et al., 2021; Xie et al., 2020) and gamma delta (γδ) T cells (Rozenbaum et al., 2020). Gamma-delta (γδ) T cells are relatively poorly investigated compared with αβ T cells as an adoptive transfer therapy. But lately, they have been under the spotlight for novel T-cell therapeutics and cancer immunotherapy research (Park & Lee, 2021; Rozenbaum et al., 2020). A primary reason for that is their fascinating combination of properties and functions (Chien et al., 2014; Kabelitz et al., 2020).

## Gamma delta (γδ ) T cells

In 1985, the γδ  T cell receptor was cloned for the first time. This identifies cells as a T cell subset and leads to initial insights into their functions (Hayday et al., 1985; Lanier et al., 1985; Saito et al., 1984). Initial research on γδ T cells focused on understanding the mechanisms behind their activation and investigating the gamma delta (γδ )receptors' methods of recognition and signalling pathways. At first sight, there were significant assumptions on similarities between αβ T cells and γδ T cells activation mechanisms and functionalities. Early studies even emphasised the ability of γδ T cells to recognise and respond to MHC molecules, an essential step for αβ T cell activation and recognition of foreign antigens (Weintraub & Hedrick, 1995). γδ T cells represent only 5% of the cells in lymphoid organs and 1-10% of CD3-positive T cells. The T cell receptor of γδ T cells is composed of gamma and delta subunits, which are made by the recombination of variable (V), diversity (D), and joining (J) and constant (C) regions of the

receptor. Moreover, in contrast to αβ T cells, the receptor repertoire of γδ T cells is minimal, mainly due to the restricted usage of V gamma and V delta gene segments.

Additionally, γδ T cell presence varies among different age groups, species, and patients. Reports show that rodents have higher numbers of γδ T cells than humans. Vγ9Vδ2 T cells are the predominant subtype of γδ T cells in the blood (Vantourout & Hayday, 2013; Weintraub & Hedrick, 1995).

## Unique characteristics of γδ T cells

Several characteristics make γδ T cells very promising for their use in therapies. Firstly, is their ability to bridge between innate and adaptive immunity and have essential roles in stress surveillance and tumour immunity; γδ T cells express receptors of innate immunity (i.e., NKG2D receptor (killer cell lectin-like receptor subfamily K, member 1 (Klrk1)) of NK cells and γδ T cell receptor), allowing them to be activated without the need for antigen processing. Secondly, how T cells are activated after recognising foreign antigens is similar to adaptive immune cells. Thirdly, similar to αβ T cells, γδ T cells can effectively kill and lyse a wide range of tumours and leukemic cells. In addition, they can differentiate into multiple types of T helper cells and secrete various cytokines with a range of functions, such as IFN-γ, IL-4 (Interleukin-4), IL-10 (Interleukin-10) and IL-17 (Interleukin-17) (Chien et al., 2014). A distinct feature of γδ T cells is the relative lack of expression of CD4 and CD8 molecules, consistent with their recognition of foreign antigens independent of MHC/HLA molecules. Similar to αβ T cells, γδ T cells have a CD3 signalling complex; they also share homologies between the α (from αβ T cells) and γ chains (from γδ T cells) subunits, as well as homologies between the light chains of the γδ T cell receptor and immunoglobulin receptors (Chien et al., 2014; Weintraub & Hedrick, 1995).

Recent publications have expanded on the possible functions of γδ T cells. For example, a publication reported that bisphosphonates from the cholesterol synthesis pathway have immunomodulatory roles in activating γδ T cells. The study found that the isoprenoid precursor, isopentenyl diphosphate (IPP), produced in the mevalonate pathway of cholesterol synthesis, must be present at very high concentrations to be recognised by Vγ9Vδ2 T cells. A dysregulated

mevalonate pathway results in high levels of accumulated IPP and stressed cells targeted by γδ T cells (Figure 1.2 (B)). In contrast, γδ T cells do not recognise healthy cells with low levels of IPP (Kabelitz et al., 2020).

Two groups also identified BTN3A1 (Butyrophilin Subfamily 3 Member A1) and the intracellular signalling domain B30.2 as central activators of γδ T cells (Sandstrom et al., 2014; Sebestyen et al., 2016). They showed that full antigen-based activation of γδ T cells depend on the presence of butyrophilin molecules located on the transmembrane of target cells. A more detailed investigation revealed that the recognition mechanism of γδ T cells is moderated by the binding of RhoB GTPase molecules (member of the Ras super-family and Rho sub-family of GTPases) to the B30.2 domain in BTN3A1 in tumours, followed by an alteration to BTN3A1 confirmation (Sebestyen et al., 2016). Karunakaran et al in 2020 identified another molecule, BTN2A1(Butyrophilin Subfamily 2 Member A1), that binds on Vγ9 regions of the γδ TCR and cooperates with BTN3A1. Following the initial recognition of phospho-antigens in tumours, BTN2A1-BTN3A1 maintains the engagement of additional TCR regions and leads to complete activation of γδ T cells (Karunakaran et al., 2020; Rigau et al., 2020).

A few studies suggest the recruitment of additional molecules, yet to be identified, to the CD3 receptor to activate γδ T cells (Karunakaran et al., 2020). Two recent publications (Karunakaran et al., 2020) hypothesise that following the binding of BTN2A1 to germline encoded region of the Vγ9 chain, binding of an identified ligand to CDR3 (complementarity determining region 3) region of the delta TCR is initiated and proposing independence on GTPase RHO molecules for γδ T cells activation (Karunakaran et al., 2020). Further elucidation and confirmation of the germline interaction of γδ TCR with BTN2A1 was shown recently by Willcox and coworkers (Willcox et al., 2023). Previous studies reported F1-ATPase (F-Type ATPase) and hMSH2 (mismatch repair genes) as potential ligands recognised by γδ T cells (Chien et al., 2014; Kabelitz et al., 2020). Novel activators of γδ T cells continue to be studied, and their mechanisms are not fully understood.

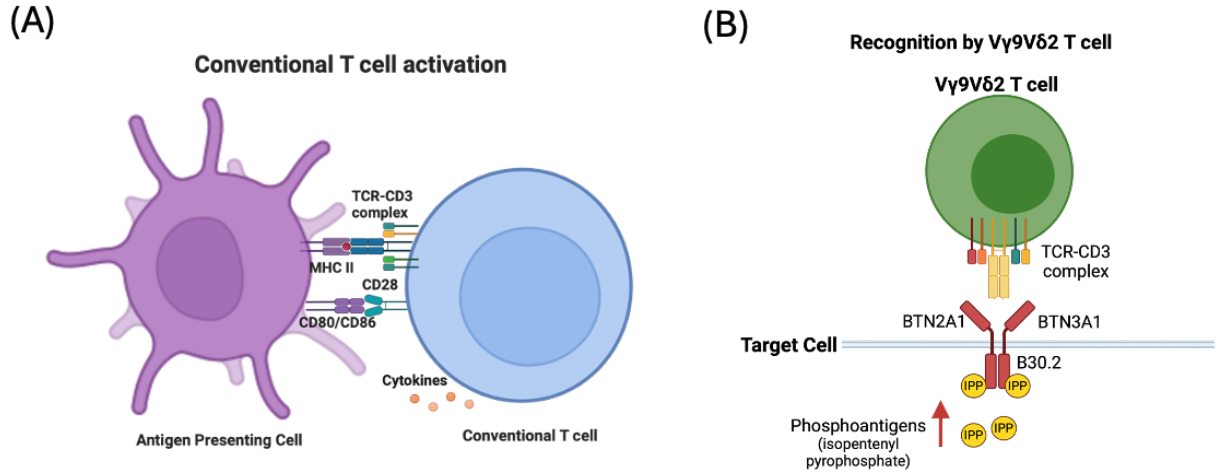



*Figure 1.2: A diagram shown Activation of (A) conventional T cells and (B) Vγ9Vδ2 T cells following recognition of antigen or target cells.*

## Synthetic Biology approaches to study exhaustion in γδ T cells.

### *CRISPR*

Bacteria use RNA-mediated adaptive defense systems called clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) proteins as protection from invading viruses and plasmids. It was discovered following a published report on identifying repetitive DNA sequences in bacteria. The mechanism starts with the initial integration of foreign short-sequence fragments of invaders into the host DNA. Then, transcription of the short fragments into CRISPR RNA leads to its binding to the foreign nucleic acids of invaders, causing its silencing with the help of Cas proteins (Brouns, 2012).

Later, viruses were also discovered to have short DNA sequences similar to what was found in CRISPR in bacteria, leading to the critical observation of the role of CRISPR in adaptive immunity against viral infections (Jinek et al., 2012).

In 2012, the first publication showcasing CRISPR as a new tool was accomplished by Jennifer A. Doudna and Emmanuelle Charpentier (Jinek et al., 2012). CRISPR has improved our understanding of the roles of genes in diseases and their effects on specific pathways of interest and sheds light on understanding the genomes of multiple organisms (i.e., humans, plants and animals). CRISPR has the potential to improve clinical outcomes of diseases that were once known as untreatable. In biology, the impact of CRISPR research has been undeniable, specifically with gene editing and developing new mouse models (J. Y. Wang & Doudna, 2023). One powerful application of CRISPR-Cas is the identification of new key targets or regulators in immunology, cell biology and cancer research using CRISPR screens (Bock et al., 2022).

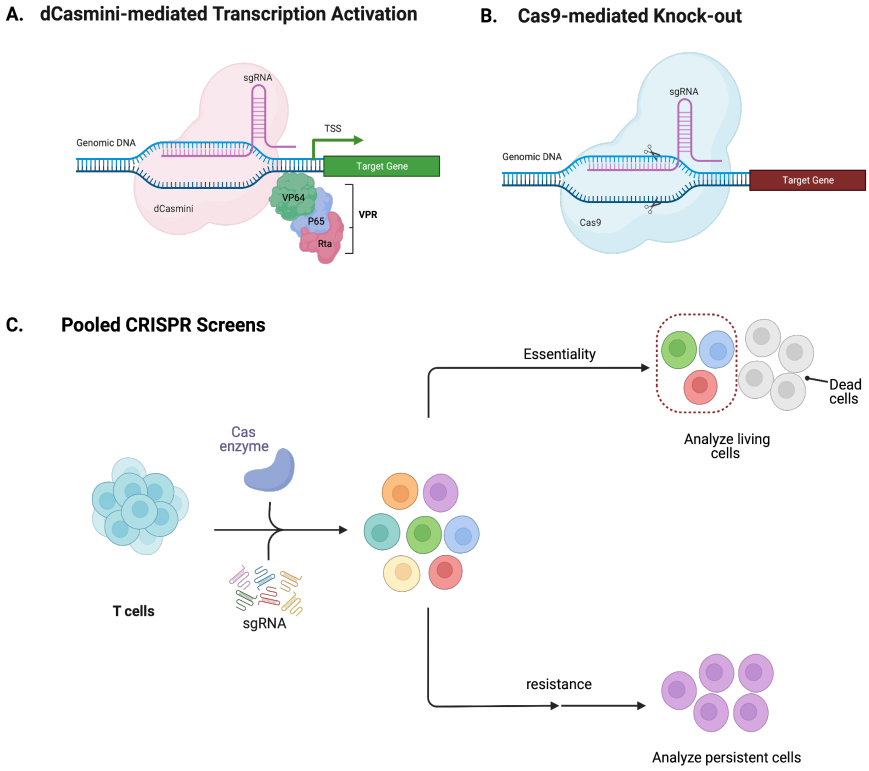



*Figure1.3: CRISPR tools and applications. A. CRISPR mediated transcriptional activation using deactivated miniature Cas12f and transcriptional activation domain VPR. B. CRISPR knock-out of a gene of interest using Cas9 enzyme. C. Application of using CRISPR in pooled sgRNA screens.*

### *CRISPR Recognition*

#### Cas9 nuclease

The most used CRISPR editor is CRISPR-cas9 from streptococcus pyogenes, a system that uses a guide RNA (gRNA or sgRNA) targeted to a sequence and a protein (Cas9) that induces a cut at the recognition site. Cas9 is a large protein with multiple domains; one is for recognition, and one for nuclease activity. The recognition domain contains PAM (protospacer adjacent motif) interacting motifs, while the nuclease domains are composed of RuvC-like and HNH-like domains (Xue & Greene, 2021).

Upon recognition of PAM, unwinding of the dsDNA is initiated, followed by R loop formation from the gRNA-DNA hybridisation when the gRNA fully complements the DNA sequence. R loop triggers a conformational change in Cas9 protein and activation of the nuclease activity, inducing a double strand break of the target DNA at three bases upstream of the PAM via the HNH-like domain. The RuvC-like nuclease domain cut is done on the non-targeted DNA strand and 3-5 base pairs upstream of the PAM sequence (Xue & Greene, 2021).

#### Guide RNAs

Guide RNAs are used to guide the Cas protein to targeted sites. Different DNA sequences can be easily recognised by designing different gRNAs (J. Y. Wang & Doudna, 2023). gRNAs can be produced by transcription from viral plasmids or synthesised chemically. Synthesised gRNAs can be modified to prevent their digestion in cells and increase their stability to enhance target recognition; they can also be subject to chemical backbone modifications to avoid immunostimulatory activity and cell toxicity by single-strand recognition via Toll-like receptor 7 (Xue & Greene, 2021).

### *CRISPR-induced gene knock-outs*

Since the discovery of CRISPR, many tools and applications have been generated and tested. CRISPR knock-outs are generated mainly using the CRISPR-Cas9 ribonucleoprotein and a single guide RNA to target the Cas9 enzyme to a specific location. Upon targeting a sequence, the Cas9 induces a double-strand break (DSB) in the DNA. In eukaryotes, DSBs are naturally repaired by non-homologous end joining or homology-directed repair.

Non-homologous end joining (NHEJ) is the preferred method by cells to repair DSBs, but it usually introduces errors and mutations. NHEJ involves the ligation of two ends of the DSB with minimal processing, resulting in low accuracy. Homology-directed repair (HDR) can use a double-stranded DNA or single-stranded template for repair. HDR is accurate and more desirable for gene editing approaches (Xue & Greene, 2021). CRISPR potential for knock-out of genes has been validated as an easy and fast approach to studying gene functions and generating novel animal models.

The latest CRISPR applications have introduced point mutations like base editing or prime editing to introduce small insertions and deletions with point mutations, enabling saturation genome editing. Base editors are Cas9 nickase proteins fused to a deaminase domain to convert one DNA base to another without causing double-strand breaks. Base editing expands the potential of point mutation corrections and editing without dependency on double-strand breaks, DNA repair pathways, or DNA templates (Komor et al., 2016). Prime editing can introduce and delete sequences without causing DSB. A prime editor consists of several Cas9 proteins fused to reverse transcriptase and has a gRNA that directs it to the target to act as a template for the required edited sequence (J. Y. Wang & Doudna, 2023).

### *CRISPR screens*

Following the success of single-gene knock-outs and the advancement of CRISPR technology throughout the last decade, approaches involving the parallel systematic knockout of multi-gene targets at once were implemented using CRISPR screens. CRISPR screening significantly impacts our understanding of the relationship between different gene network functions and their pathways. Crispr Screens are very valuable in identifying new targets for developing novel therapeutics. In a screen, perturbations are introduced to one or multiple genes, followed by assays to read out the functions or impacts of these genes in a cellular system. Assessment of effects on cell functions by single gene knock-outs or multi-knockouts in pooled screens is possible with advanced CRISPR technologies.

The guide RNA component of the CRISPR system is very easy to design and highly accessible. To this date, gRNA designs are available for almost every gene in the human genome. Other CRISPR tools include the use of CRISPR interference (CRISPRi) to cause the inhibition of gene targets and, using CRISPR activation (CRISPRa) to induce the activation of selected targets (Bendixen et al., 2023).

CRISPRi uses a deactivated Cas enzyme (Dead or dCas) binding downstream from a promoter binding site, causing a steric hindrance for the RNA polymerase and inhibiting transcription to the targeted gene. CRISPRa uses fusion of activation domains to Cas enzyme C-terminus. A known method for transcriptional activation is the use of a transcription factor activation domain or a hybrid of domains such as the VPR system (complex consisting of VP64, p56 and Rta activation domain), which, upon binding to DNA, activates a targeted gene (Figure 1.3A and 1.3B) (Bendixen et al., 2023; Chavez et al., 2015).

The diversity of CRISPR editing tools permits researchers to ask broad scientific questions. CRISPR has been widely used in humans, animals and plants, and its outcomes have impacted our studies of diseases such as Cancer and immunotherapy. CRISPR is one approach that led to the discovery of new regulatory elements and our increased knowledge of cell development and gene functions.

CRISPR experiments can be custom-designed for desired readouts. A few standard experiments are the study of cell proliferation or killing, screening of cell phenotypes via fluorescence-activated cell sorting (FACS) and detecting surface marker expression. Phenotyping cells following CRISPR editing can provide insights into cell development, functions and differentiation trajectories. Assays with CRISPR-edited cells can suggest the response of specific cell types against tumours or the effects of certain edits on immunity. More specifically, CRISPR editing has shown to be very promising in immunotherapy and the generation of long-persisting and better-targeting immune cells against various infections (J. Y. Wang & Doudna, 2023)

Combining CRISPR with sequencing technologies provides more depth to different systems and tracking of cellular genetic reprogramming of pathways following single or multiple edits. The discovery of new nucleases and the enhancement of the features of current Cas enzymes will contribute to the increased potential of multiplex screen developments and answering complex scientific questions (Datlinger et al., 2017).

### *Miniature Cas protein (Casmini)*

Multiple systems of CRISPR-cas have been developed and used for genome editing. The well-known SpCas9 (derived from streptococcus pyogenes) is well-documented for different applications. Another CRISPR system designed is Cas12a (derived from the Lachnospiraceae bacterium), which has also proven efficient. The dead Cas (dCas) is mutated at the nuclease domain to prevent cutting and instead fused to a transcriptional regulator (activator or inhibitor). dCas transcriptional and epigenetic regulators are widely used for different tools in cellular reprogramming; however, the sizes of their enzymes are very large and limit their potential delivery to cells (Xu et al., 2021a).

Different Cas enzymes known as smaller naturally occurring Cas effectors have been discovered, and they are known for their small and compact size. Cas12f is a known small Cas enzyme that is well-used and part of a compact class 2 V_F system of exceptionally compact RNA-guided nucleases. Cas12f is derived from uncultivated archaea and is yet to be proven for its activity in mammalian cells (Xu et al., 2021a).

In a recent report, an engineered version of Cas12f with only 529 amino acids (the smallest among Cas enzymes) has been engineered to function in mammalian cells and is called Casmini. A version of Casmini has been mutated to serve as a transcriptional activator (dCasmini) that was shown to work effectively for transcriptional activation when fused to a transcriptional activator hybrid domain (VPR). The Casmini and dCasmini enzymes are easier to use for cellular delivery and can provide enhanced efficacy due to their compact size and improved functionality (Xu et al., 2021a).

## Precision and accuracy in CRISPR technologies

With the progress achieved so far with CRISPR technologies, there are still a few challenges that we need to overcome.

First is the editing accuracy and specificity to target sites. As current tools fail to eliminate off-target effects, new designs using various nucleases have been generated, such as high-fidelity Cas enzymes like SpCas9-HF1 (Kleinstiver et al., 2016) and HiFiCas9 (Vakulskas et al., 2018). gRNAs have also been improved to reduce off-target binding (i.e., sgDesigner (Hiranniramol et al., 2020) and E-Crisp (Heigwer et al., 2014)). They are validated in the clinic with negligible off-target sites. The second is editing precision and getting the expected outcomes of using CRISPR. Precise editing is harder to control, and we still cannot anticipate effects following the double-strand break induced by Cas enzymes.

Machine learning tools are being developed to predict editing efficiencies, but they still must be fully validated for their efficacy in vivo. Competition between NHEJ and HDR repair pathways following a DSB is another source of reduced precision, as when NHEJ is the method of repair, indels are inserted. NHEJ is the non-desirable outcome for therapeutics requiring higher editing precision (J. Y. Wang & Doudna, 2023).

Understanding DNA repair pathways and how to manipulate them is necessary to enhance editing precision. One promising future editing tool is prime editing; its independence from requiring a DSB formation has proven to reduce off-target edits and shown better precision than HDR in some settings. However, prime editing still needs improvements to enhance its efficiency against other tools.  CRISPR-associated transposition is a novel tool that has only been validated in prokaryotes; it uses RNA-guided DNA transposition to enable the insertion of large DNA via transposons. Developing CRISPR-associated transposition to mammalian systems might be another way to increase precision for future applications (J. Y. Wang & Doudna, 2023).

## CRISPR off-target effects

One of the limitations of using gene editing tools like the CRISPR/ Cas 9 system is off-the-target effects. When Cas9 cuts an untargeted DNA strand, it leads to off-target effects that have potential to cause unwanted edits to the genome.

Some methods used to minimise off-target effects include *in silico* and experimental tools. *In silico* tools, such as CCTop (Stemmer et al., 2015), CasOT (Xiao et al., 2014), and DeepCRISPR (Chuai et al., 2018), are online-based platforms that can predict off-target effects caused by sgRNA sequences. However, these methods need further experimental validation.

Experimental methods include cell-free methods that operate on extracted DNA or chromatin and detect cleavages in them. Digenome-seq is a standard method used. Following the DNA editing, the analysis is performed with whole genome sequencing to detect the double-strand break (DSB) loci. However, it requires high sequencing coverage (~400–500 million reads for the human genome) due to the high background of non-specific DSB in the DNA (Kim et al., 2015). CIRCLE-seq is another method that does not require such high sequencing depth. It involves shearing and circulation of the genomic DNA by intramolecular ligation. The circular DNA is selectively linearised in this method upon Cas 9 cleavage. The non-specific linear DNA and undigested circular DNA are removed, resulting in a low background. CIRCLE-seq requires only 4-5million reads (Kim et al., 2019; Tsai et al., 2017)

## CRISPR Delivery

Various delivery methods are currently being used and improved for CRISPR systems. Delivery is dependent on the nature of CRISPR components. Physical delivery is usually achieved by electroporation or injection of CRISPR components into desired cells; it can deliver DNA plasmid, mRNA (messenger RNA) or RNP (Ribonucleoprotein) format of Cas with its gRNA but is limited to ex-vivo delivery only. Electroporation also leads to a transient activity and thus lowers the risk of off-targets. One approach involving viral delivery is when the plasmid DNA template for CRISPR HDR editing or template for Cas or guides is delivered using AAV (Adeno-associated virus), AdVs (Adenoviruses) or Lentiviruses, and this method has been used

for ex vivo and in vivo studies; AAV delivery has even been validated in clinical trials. However, as AAVs have low packaging capacity, lentiviruses offer higher packaging capacities and are widely used for editing. Lentiviral delivery causes permanent insertions of DNA in the genome and permanent expression of Cas enzyme and gRNA, increasing the risk of off-targets. Synthetic material-based delivery is when nanoparticle and polymer or peptide delivery are used to deliver any form of CRISPR systems. Although Synthetic material-based delivery offer flexible and controlled delivery, their efficiency is lower than viral delivery systems and requires many optimisations (J. Y. Wang & Doudna, 2023).

In summary, CRISPR-based tools have proven their potential in research and clinically. There are 8 Food and Drug Administration (FDA) -approved sickle cell disease trials, and the first FDA approval of treatment with CRISPR-based therapy was recently granted in November 2023 (Sheridan, 2023). To transform CRISPR into a standard practice in clinics, addressing challenges such as manufacturing difficulties and the high costs of treatments is essential. Finally, expanding the CRISPR toolbox by discovering new variants or designing improved tools would significantly enhance our future applications and accommodation of this tool in research and the clinic.

## Aims and Objectives

This work's originality lies in studying gamma delta T cells' behaviour under chronic activation and investigating crucial questions regarding their specific phenotype and functionality. Recent reports on exhaustion have focused on αβT cells whilst relatively little attention has been paid to γδ T cell exhaustion. There is an urgent need to understand how these T cells respond to chronic infections and whether they develop an exhausted phenotype.

We hypothesise that γδ T cells develop exhaustion or hyporesponsiveness, which is detected as hypoproliferation, or increased expression of exhaustion markers following chronic stimulation.

We will investigate how distinct or similar γδ T cell exhaustion is compared to αβ T cell exhaustion and identify whether exhaustion in γδ T cells associate with any unique characteristics not observed in exhausted αβ T cells.

Moreover, high-resolution technologies will be used to confirm our findings and to observe, in greater detail, potential developmental changes occurring at a transcription or an epigenetic level during chronic stimulation. There is a deep curiosity in the scientific field to learn about the origins of exhausted T cells and the development of the exhaustion lineage, which is essential for validating or negating the proposed hypothesis regarding the emergence of exhausted T cells either from biological switching of previous cells lineages or their evolution as a unique lineage.

Finally, we will evaluate the effects of knocking out putative exhaustion-promoting genes or the activation and overexpression of anti-exhaustion genes on the partial or complete recovery of exhausted T cells is crucial.

# Chapter 2

## Materials and Methods

### Isolation of primary human peripheral blood mononuclear cells

Blood buffy coats are obtained from donors in the NHS facility. Blood is first diluted with PBS (Phosphate-buffered saline) to reach a 30ml solution, then layered gently on top of 20ml Lymphoprep (AXIS-SHIELD). After that, the blood/ Lymphoprep mixture is separated by low-density centrifugation at 2400rpm and for 20min with no centrifuge braking. The blood mixture is separated into four layers following the centrifugation, with a plasma layer on top. The PBMCs (Peripheral blood mononuclear cells) are positioned at the 2nd layer, followed by the Lymphoprep solution and the granulocytes and erythrocytes at the bottom. The PBMCs layer contains the lymphocytes and monocytes and is collected, washed and counted using trypan blue (dead cells staining solution). The required amount of PBMCs is then usied to allow for subsequent expansion and activation. Around 3 x $10^5$ cells/well are seeded in 96 well plates with complete RPMI-1640 medium with L- glutamine (Gibco), penicillin (100 IU/ml, Sigma-Aldrich), streptomycin (100 µg/ml, Sigma-Aldrich) and 10% FCS (v/v, (Gibco).

### T cells activation and phenotyping

#### *Chronic stimulation with OKT3*

Around 3 x $10^5$ cells/well are plated in a 96-well plate with OKT3 (Anti-CD3 monoclonal, clone OKT3) (1ug/ml) and IL-2 (2600 IU/ml). Additional OKT3 (1ug/ml) was provided to activate the cells on days 3 and 6. IL-2 was replenished with new media every 2-3 days. Cells were analysed on days 0, 6, 9 and 15 for their phenotype after antibody staining.

### *Chronic stimulation with Zol and OKT3*

Around 3 x $10^5$ cells/well are plated in 96 well plates with 5µM Zoledronic acid (Zol) (Actavis) and IL-2 (2600 IU/ml, Proleukin). IL-2 was replenished with new media every 2-3 days. Then, OKT3 (1ug/ml, BioLegend) was used to further activate the cells on days 3 and 6. Cells were analysed on days 0,3, 6 and 9 for their phenotype after antibody staining.

### *Chronic stimulation with Daudi and anti-BTN3A (20.1)*

Daudi cells were first pre-coated with 1µg anti-BTN3A (20.1) antibody in 100ul for 1 hr in 37C. Then, the cells were washed and plated with γδT cells at an effector-to-target ratio of 10:1 with IL-2 (2600 IU/ml, Proleukin) in 96-well plates. IL-2 was replenished with new media every 2-3 days. Then, anti-BTN3A (20.1) pre-coated Daudi cells are also used to further activate the cells on days 3 and 6. Cells were analysed on days 0,3, 6 and 9 for their phenotype after antibody staining.

## Staining

### *Surface Markers*

Before staining, cells were collected and washed by centrifugation at 300 relative centrifugal force (rcf) for 5 min. Cells were then resuspended in 50µl PBS followed by the addition of human FC receptor block solution (Biolegend) and staining with a dead cell stain kit (ThermoFischer Scientific). After mixing the samples well by vortex, they were incubated for 15 min, protected from light and at 4°C, then washed with PBS by centrifugation at 300 rcf for 5 min. Surface markers antibodies were then added, mixed well by a 30-second vortex and Incubated for 20 min (4°C  protected from light) and washed with cell staining buffer (BioLegend).
Finally, samples were fixed and permeabilised using a transcription factor staining buffer kit (Foxp3 kit from ThermoFisher Scientific) for subsequent intracellular staining. For fixing and permeabilising the cells, the buffer was prepared by diluting a fix/perm concentrate with three

parts of a fix/perm diluent solution. If only surface staining is required, the cells can be analysed immediately by flow cytometry or fixed with IC fixation buffer (ThermoFisher Scientific) for analysis on a different day.

### *Intracellular Markers*

For intracellular staining, the cells must be fixed after staining of surface markers and permeabilised for the intracellular staining. The FoxoP3 kit was used for this staining. First, buffer preparation is required by diluting the fix/perm concentrate with 3 parts of fix/perm diluent solution. A permeabilisation buffer was also prepared by diluting it 10X with distilled water. To start, 1ml of prepared fix/perm solution was added to each sample, mixed well and followed by incubation at 4°C for 15 min. Then, the cells were washed using the diluted perm buffer (2ml/sample) and centrifugation was done at 300 rcf for 5 min. The intracellular staining antibodies were added, mixed well by vortex for 30 seconds and incubated 30-60 min at 4°C and protected from light. Finally, the cells were washed using the perm buffer followed by centrifugation. The cells can be resuspended in perm buffer and analysed immediately or on another day (1-2 days maximum) if left at 4°C and protected from light. Leaving intracellularly stained samples for more prolonged periods could affect the results.

## E changeProliferation assay

Cells under chronic activation were investigated for their proliferative capacity. To measure proliferation, a tracking dye was used to monitor the proliferative status of cells during the time course of a study. CellTrace™ Violet (CTV) (ThermoFisher Scientific) is a well-known tracking dye that incorporates inside the cells and remains stable over time. The dilutions of the dye across a study indicate the number of cell divisions that occurred and the proliferative capacity of a specific population. For labelling cells, 1µl of CellTrace™ Violet was used per 1 ml cells (1:1000) followed by a 20-minute incubation at 37°C (protected from light). The dye is then quenched by adding 5x volume of media to the stained solution and re-incubation at 37°C for 5

minutes. Finally, the cells were washed by centrifugation (300 rcf for 5 min) and resuspended in the required amount of media to proceed with stimulation or staining.

## Cell death and apoptosis assay

Cell death was detected using Annexin V and propidium iodide (PI). Annexin V provides a sensitive signal to detect apoptotic cells and has been used widely for flow cytometric analysis. PI is commonly used as a viability stain and indicates necrotic or late apoptotic cells by losing their membrane integrity. For cell staining, the cells were washed by centrifugation after surface antibodies staining and addition of the required amount of Annexin V solution (5µl). Samples were incubated 15-20 minutes, protected from light and then washed. PI staining follows the fixation and permeabilization of cells using the transcription factor staining buffer kit (Foxp3 kit from ThermoFisher Scientific). PI (2µl) was used to stain the cells in a 50-100µl suspension. The cells were incubated, protected from light for 30 minutes, washed and analysed using flow cytometry.

## Fluorescence-activated cell sorter

Institution: University College London

Facility: Great Ormond Street Institute of Child Health

Cytometer: LSR II (355/405/488/633)

*Table 2.2: Instrument configuration*

| **LASER (NM)** | **Band Pass** | **Long Pass** | **Optimal Colours** | |
|---|---|---|---|---|
| LASER | ***355*** | ***450 / 50*** | -- | **DAPI**<br>**LIVE/DEAD™ Fixable Blue** |

| | | | | |
|---|---|---|---|---|
| | | | | **DyLight 350** |
| | | ***530 / 30*** | ***505*** | |
| LASER | ***405*** | ***450 / 50*** | -- | **Pacific Blue**<br>**eFluor 450**<br>**VioBlue** |
| | | ***525 / 50*** | ***475*** | **Brilliant Violet 510** |
| | | ***560 / 20*** | ***545*** | **Brilliant Violet 510**<br>**Brilliant Violet 570** |
| | | ***605 / 12*** | ***595*** | **Brilliant Violet 605** |
| | | ***655 / 8*** | ***630*** | **Brilliant Violet 650** |
| | | ***710 / 50*** | ***685*** | **Brilliant Violet 711** |
| LASER | ***488*** | ***530 / 30*** | ***505*** | **Alexa Fluor 488** |
| | | ***575 / 26*** | ***550*** | **PE** |
| | | ***610 / 20*** | ***600*** | **PE-Texas Red**<br>**PI**<br>**PE-Dazzle 594** |

| | | | | |
|---|---|---|---|---|
| | | *660 / 20* | *655* | **PE-Cy5** |
| | | *695 / 40* | *685* | **PerCP-Cy5.5** |
| | | *780 / 60* | *755* | **PE-Cy7**<br>**PE-Vio770** |
| LASER | *633* | *660 / 20* | -- | **Alexa Fluor 647**<br>**APC** |
| | | *730 / 45* | *710* | **Alexa Fluor 700** |
| | | *780 / 60* | *755* | **APC-Cy7**<br>**APC-Vio770** |

## List of Panels

### *OKT3 Model*

*Table 3.2: FACS Panel I- Surface markers*

| Laser | EM Filter | Marker | Colour / Format | Host / Target | Isotype | Clone | Company | Catalogue |
|---|---|---|---|---|---|---|---|---|
| 355 | 450/5 | Live/Dead fixable dead cell stain kit | | anti-All Species | | | ThermoFisher | L23105 |
| 488 | 610/20 | CD3 | PE-Dazzle 594 | Mouse anti-Human | IgG2a κ | HIT3a | BioLegend 877-246-5343 | 300335 |
| 488 | 780/60 | TCR alpha/beta | PE-Vio770 | Mouse anti-Human | IgG2b | BW242/412 | Miltenyi Biotec 866-811-4466 | 130-113-532 |
| 405 | 450/50 | TCR V delta 2 | VioBlue | Mouse anti-Human | IgG1 κ | 123R3 | Miltenyi Biotec 866-811-4466 | 130-101-152 |
| 633 | 730/45 | TIM-3 | DyLight 680 | Mouse anti-Human | IgG1 κ | F38-2E2 | Novus Biologicals 888-506-6887 | NBP2-27221FR |
| 405 | 710/50 | Lag-3 | Brilliant Violet 711 | Human | | | Custom reagent - | |

*Table 2.4: FACS Panel II – Cell death and apoptosis*

| Laser | EM Filter | Marker | Colour / Format | Host / Target | Isotype | Clone | Company | Catalogue |
|---|---|---|---|---|---|---|---|---|
| 633 | 660/20 | CD3 | APC | Mouse anti-Human | IgG2a κ | BW264/56 | Miltenyi Biotec 866-811-4466 | 130-113-125 |
| 633 | 780/60 | VD2 | APC-Vio770 | Human | | | Custom reagent - | |
| 488 | 660/20 | PI | PI | N/A anti-All Species | N/A | N/A | Miltenyi Biotec 866-811-4466 | 130-093-233 |
| 405 | 710/50 | Annexin V Apoptosis Kit | Brilliant Violet 711 | N/A anti-Human | N/A | N/A | BD Biosciences 877.232.8995 | 563972 |

### *Zol/OKT3 model*

*Table 2.5: Panel I – Surface markers*

| Laser | EM Filter | Marker | Colour / Format | Host / Target | Isotype | Clone | Company | Catalogue |
|---|---|---|---|---|---|---|---|---|
| 355 | 450/50 | Live/Dead fixable dead cell stain kit | | anti-All Species | | | Thermo Fisher Scientific | L23105 |
| 405 | 450/50 | CellTrace Violet | CellTrace Violet | N/A anti-All Species | N/A | N/A | Thermo Fisher Scientific 1 800 955 6288 | C34557 |
| 633 | 780/60 755LP | TCR V delta 2 | APC-Vio770 | cell line anti-Human | IgG1 | REA591 | Miltenyi Biotec 866-811-4466 | 130-113-509 |
| 488 | 780/60 755LP | TCR alpha/beta | PE-Vio770 | Mouse anti-Human | IgG2b | BW242/412 | Miltenyi Biotec 866-811-4466 | 130-113-532 |
| 488 | 575/26 550LP | Lag-3 | PE | cell line anti-Human | IgG1 | REA351 | Miltenyi Biotec 866-811-4466 | 130-120-610 |
| 633 | 730/45 710LP | TIM-3 | Alexa Fluor 700 | Mouse anti-Human | IgG1 κ | F38-2E2 | Thermo Fisher Scientific 1 800 955 6288 | 56-3109-42 |

*Table 2.6: Panel II – Intracellular transcription factors*

| Laser | EM Filter | Marker | Colour / Format | Host / Target | Isotype | Clone | Company | Catalogue |
|---|---|---|---|---|---|---|---|---|
| 355 | 450/50 | Live/Dead fixable dead cell stain kit | | anti-All Species | | | Thermo Fisher Scientific | L23105 |
| 405 | 450/50 | CellTrace Violet | CellTrace Violet | N/A anti-All Species | N/A | N/A | Thermo Fisher Scientific 1 800 955 6288 | C34557 |
| 633 | 780/60 755LP | TCR V delta 2 | APC-Vio770 | cell line anti-Human | IgG1 | REA591 | Miltenyi Biotec 866-811-4466 | 130-113-509 |
| 488 | 780/60 755LP | TCR alpha/beta | PE-Vio770 | Mouse anti-Human | IgG2b | BW242/412 | Miltenyi Biotec 866-811-4466 | 130-113-532 |
| 488 | 575/26 550LP | TOX | PE | cell line anti-Human | IgG1 | REA473 | Miltenyi Biotec 866-811-4466 | 130-120-716 |
| 633 | 660/20 | TCF7 | Alexa Fluor 647 | Mouse anti-Human | IgG1 κ | 7F11A10 | BioLegend 877-246-5343 | 655204 |

*Table 2.7: Panel III – Cell proliferation tracking*

| Laser | EM Filter | Marker | Colour / Format | Host / Target | Isotype | Clone | Company | Catalogue |
|---|---|---|---|---|---|---|---|---|
| 355 | 450/50 | CellTrace Violet | CellTrace Violet | N/A anti-All Species | N/A | N/A | Thermo Fisher Scientific 1 800 955 6288 | C34557 |
| 633 | 660/20 | CD3 | APC | Mouse anti-Human | IgG1 κ | SK7 | BioLegend 877-246-5343 | 344811 |
| 633 | 730/45 | TCR V delta 2 | APC-Cy7 | Mouse anti-Human | IgG1 κ | B6 | BioLegend 877-246-5343 | 331440 |

## Functional Assays

### *Co-culture using Daudi and Vδ2 T cells*

PBMCs were isolated on day 0 and activated with Zol (34ul/ml) +IL-2 (2600IU/ml). On Day 3, media +IL-2 was replaced for ZOL only condition and Media with IL-2 and OKT3 (1µg/ml) was added for the second condition (ZOL + OKT3). Daudi cells were thawed and cultured for a few days before being used in the assay. On day 4, the cells were plated in a U-bottomed 96-well plate with different ratios of Vδ2 T cells to Daudi cells. The ratios used were 1:1, 1:5 and 1:10 Vδ2 T cells to Daudi cells. All samples were analysed in triplicates and for three donors. Cell trace labelling was performed on day 4 on both cell types to allow efficient tracking and functionality analysis. Cell trace violet was used for Vδ2 T cells labelling, and Cell trace yellow was used for Daudi cells. Analysis was performed after 24 and 48 hours of the co-culture. Counting beads (Precision Count Beads™) from Biolegend (424902) were added for each well 5µl per well (5300 beads/well). The panel used for flow cytometry is shown in Table 2.7.

Another co-culture with the exact specification as the day 4 culture was set on day 5 to examine if there were any differences following two days of OKT3 stimulation in correspondence to 1 day after OKT3 stimulation. The cells were also labelled for cell trace violet (Vδ2 T cells) and cell trace yellow (Daudi). Analysis was performed after 24 and 48 hours of the co-culture. Counting beads (Precision Count Beads™) from Biolegend (424902) were added for each well 5ul per well (5300 beads/well). The panel used for flow cytometry is shown in Table 2.7.

### *3.1.1. Cell Trace Staining Protocol*

For cell trace staining, instruction from the manufacturer is followed. Briefly, Cell Trace was resuspended in DMSO (Dimethylsulfoxide) and then diluted 1000 times in warm PBS (at 37°C) for 5 µM staining solution. The cells were collected by centrifugation and spun at 300xg. The supernatant is removed, and the pellet was resuspended in a 5ml staining solution (CellTrace Violet for Vδ2 T cells and CellTrace Yellow for Daudi). After resuspending, the cells were incubated at 37°C for 20 min and kept in the dark. After the incubation, 5x media was added to the cells, incubated for 5 min, then collected by centrifugation (5 min at 300 xg) and resuspended in the required media for setting the culture.

*Table 2.7: Flow cytometry panel of Vδ2 T cells and Daudi co-culture*

| Laser | EM Filter | Marker | Colour / Format |
|---|---|---|---|
| **405** | 450/45 | Vδ2 using CellTrace Violet | CellTrace Violet |
| **561** | 585/42 | Daudi using CellTrace yellow | CellTrace yellow |
| **488** | 525/40 | CD3 | FITC |
| **638** | 780/60 | Viability | APC vio 770 |

*CFSE tracking in Vδ2 T cells.*

PBMCs were isolated on day 0 and activated with 5µM Zoledronic acid (Actavis) and IL-2 (2600 IU/ml, Proleukin). On Day 3, media +IL-2 was replaced for ZOL only condition and Media with IL-2 and OKT3 (1ug/ml, BioLegend) was added for the second condition (ZOL + OKT3). The cells were then stained for cell tracking across multiple days. CFSE stain (intense fluorescent staining) was resuspended in DMSO and then diluted 1000 times in warm PBS (at 37°C) for 5 µM staining solution. The cells were then incubated with CFSE at 37°C for 20 min in the dark. After the incubation, 5x media was added to the cells, incubated for 5 min, then collected by centrifugation (5 min at 300 xg) and resuspended in the required media for setting the culture. The first time-point was day 4; Vδ2 T cells were stained with CellTrace yellow and cultured for three days while tracking of their proliferation was performed daily. On day 5, the cells were stained for each day and followed up to day 7 (Figure 2.1). Counting beads (Precision Count Beads™) from Biolegend (424902) were added for each well 5ul per well (5300 beads/well). Analysis was performed using flow cytometry with a panel shown in Table 2.8. Cell proliferation was analysed to compare the ZOL only and ZOL+OKT3 conditions and identify any differences and the emergence of proliferative populations across the time points studied. The flow cytometry panel is shown in Table 2.8.

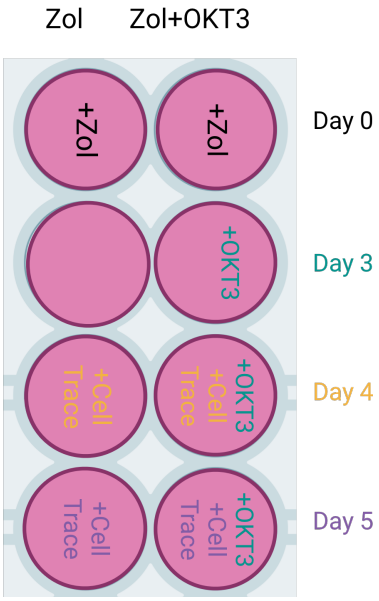



*Figure 2.1: Schematic of culture set-up for cell tracking using CFSE. The two different conditions of Vδ2 T cells stimulation is shown (Zol for single stimulation and Zol+OKT3 for double stimulation).*

*Table 2.8: Flow cytometry panel of Vδ2 T cells Cell Tracking using CellTrace*

| Laser | EM Filter | Marker | Colour / Format |
|---|---|---|---|
| **405** | 450/45 | Vδ2 | VioBlue |
| **561** | 585/42 | proliferation using CellTrace yellow | CellTrace yellow |
| **405** | 610/20 | PD-1 | BV605 |
| **561** | 780/60 | Lag-3 | PE-dazzle 594 |
| **355** | 450/50 | viability | UV |

### *Re-stimulation assay of Vδ2 T cells with target cells*

PBMCs were isolated on day 0 and activated as described previously. On Day 3, media +IL-2 was replaced for the ZOL-only condition, and Media with IL-2 and OKT3 (1µg/ml) was added for the second condition (ZOL + OKT3). Target cells (Daudi, Jurkat and THP-1) were thawed and cultured for a few days before being used in the assay. On day 4, the cells were plated in a U-bottomed 96-well plate with a ratio of 1:1 Vδ2 T cells to target cells. Conditions were for 1x, 2x and 3x stimulation with target cells. The first target cell addition was on day 4, the second was on day 5, and the third was on day 6. All samples were performed done in triplicates and for three donors. Counting beads (Precision Count Beads™) from Biolegend (424902) were added for each well 5ul per well (5300 beads/well). Analysis was performed using flow cytometry on day 7 with the panel in Table 2.9.

*Table 2.9: Flow cytometry panel of Vδ2 T cells re-stimulation assay using different target cells*

| Laser | EM Filter | Marker | Colour / Format |
|---|---|---|---|
| **405** | 450/45 | Vδ2 | VioBlue |
| **405** | 610/20 | PD-1 | BV605 |
| **561** | 610/20 | Lag-3 | PE-dazzle 594 |
| **355** | 450/50 | viability | UV |

## Transcriptomic analysis of Vγ9δ2 T cells

### *Bulk-RNA-seq 1 (Experiment 1 (n=7))*

Three donors of ZOL expanded gamma delta T cells and four donors for double-stimulated (ZOL and OKT3) cells were used in the first study. The cells were all processed by positive selection of Vδ2 T cells in the FACS sort machine to ensure that Vδ2 T cells resemble more than 90% of the bulk RNA extracted and used for this study. At least one hundred thousand cells were collected per sample.

The cells were then sent as pellets to the genomics facility at UCL for RNA extraction and library preparation (Figure 2.2). Initial processing of sample reads and quality control results, in addition to data analysis for differentially expressed genes, were processed by UCL genomics.

Comparisons were made between the double-stimulated sample (OKT3 in the analysis) and the control samples (ZOL). For sequencing, 16 million reads were used per sample.

### *Bulk-RNA-seq 2 (Experiment 2 (n=8))*

The same experiment was repeated to confirm the findings, and additional donors were used. For this, eight total donors were used, four for control samples (ZOL) and 4 for double-stimulated samples (OKT3).

The cells were all processed by positive selection of Vδ2 T cells in FACS Aria III sort machine to ensure that Vδ2  T cells resemble more than 90% of the bulk RNA extracted and used for this study. At least one million cells were collected per sample.

For the second experiment, the cells were then sent as pellets to the Genewiz company for RNA extraction and library preparation. Initial processing of sample reads and quality control results, in addition to data analysis for differentially expressed genes, were processed by Genewiz in addition to Full analysis, including functional comparisons. All Comparisons were made for the differences between the double-stimulated sample (OKT3 in the analysis) and the control samples (ZOL). For sequencing, 20-30 million reads were used per sample.

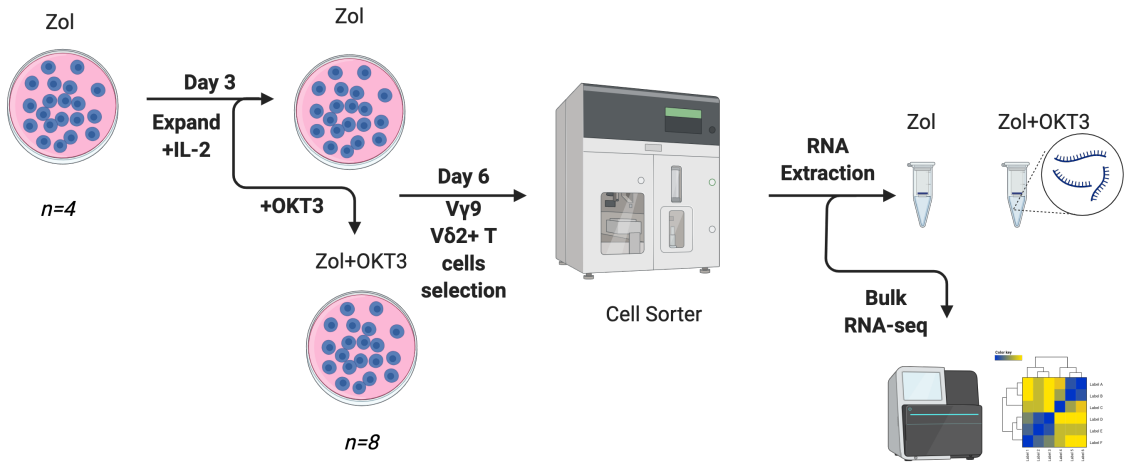



*Figure 2.2: Schematic representation of methodology of expanding and isolating chronically stimulated Vγ9Vδ2 T cells prior to RNA isolation transcriptome profiling..*

### *R analysis*

R was used for all analyses required. Packages used include ggrepel, ggplot2, reshape2, amap, BiocManager, clusterProfiler, org.Hs.eg.db, dbplyr, msigdbr, enrichplot, igraph.

The full R script is provided in the appendix.

## CRISPR Knock-out Screen on Vδ2 T cells

### *Targets selection for Knock-out screen*

Following the full transcriptomics analysis of Vδ2 T cells (single or double-stimulated) on day 6, target genes significantly enriched or depleted in the double-stimulated cells were chosen for the CRISPR screen. Along with genes differentially expressed in double-stimulated Vδ2 T cells, a few known published genes that are associated with conventional T cell exhaustion were also

included as targets to be included in the CRISPR screen. Cloned targets include TOX1, TOX2, TOX3, LAG3, TIM-3, NFATc1, BLIMP-1 and EOMES.

### *Cas9 system and gRNAs design*

Several CRISPR systems were considered for the screen. As many publications have validated the CRISPR Cas9 system, it was chosen and ordered as a vector with Cas9 enzyme, a region to insert selected gRNAs through cloning and with a green fluorescent protein (GFP) pLentiCRISPR-EGFP was a gift from Beat Bornhauser (Addgene plasmid # 75159; http://n2t.net/addgene:75159; RRID: Addgene_75159) (McComb et al., 2016)

CRISPick software was used to design all gRNAs for chosen targets. Five gRNAs were designed per target and selected by the highest knock-out scores and low off-target effects as suggested by the software algorithm.

TWIST bioscience was chosen to produce the library and clone the gRNAs into the vector, as they have expertise in CRISPR library designs.

### *Generation of Lentiviral vector*

Required plasmids for virus production include packaging plasmids (Gag, Pol and Rev), the expression vector (CRISPR screen vector) and the envelope plasmid (RdPro). Plasmids were transfected into LentiX-293T cells (from TAKARA) using GeneJuice (Merck Millipore). LentiX-293T cells were plated in 100mmx20mm culture plates at 80-90% confluency on transfection day, usually 5e6 cells the day before transfection or 2.5e6 cells two days before transfection. On transfection day, plasmids were mixed at the required concentration. GeneJuice was added to OptiMEM media, vortexed and incubated for 5 minutes at RT. Plasmid mix was added to GeneJuice and OptiMEM, slowly mixed by pipetting and incubated for 15 minutes. Finally, 500ul of the mix was added to each plate of LentiX-293T cells dropwise, rocked upward-downward and sideways, then left at the incubator @37C. The volumes of each reagent per plate are shown in Table 2.10.

After 72 hours, LentiX-293T cells should be green under a fluorescent microscope, and virus-containing supernatant was harvested in 50 ml tubes. The supernatant was spun in a centrifuge to exclude dead cells (400xg for 10 min). Finally, the virus supernatant was transferred to new tubes, balanced and centrifuged overnight at 2000rpm and 4°C. The following day, the supernatant was discarded and 200µl of optimum was added to each pellet and left on ice for 20 minutes before resuspending and aliquoting in vials. The final concentration of the virus was 300X the harvested concentration.  Virus vials must be left on dry ice immediately until all aliquoting was done, then moved to -80. The thawed virus cannot be frozen and used again.

Titrations of the virus were done by plating 1e5 cells per well in a 6-well plate and making a serial dilution of the virus before adding it to the cells.

*Table 2.10: LentiX-293T cells transfection reagents volumes for 100mmx20mm plate*

| Reagent | Volume | Ratio |
|---|---|---|
| **OptiMem** | 470ul | |
| **GeneJuice** | 30ul | |
| **CrLKo** | 6.25ug | 4 |
| **Env (RdPro)** | 1.56ug | 1 |
| **Gag/pol (pRRe)** | 3.125ug | 2 |
| **Rev (pRSV)** | 1.56ug | 1 |

## *CRISPR Knock-out screening assay*

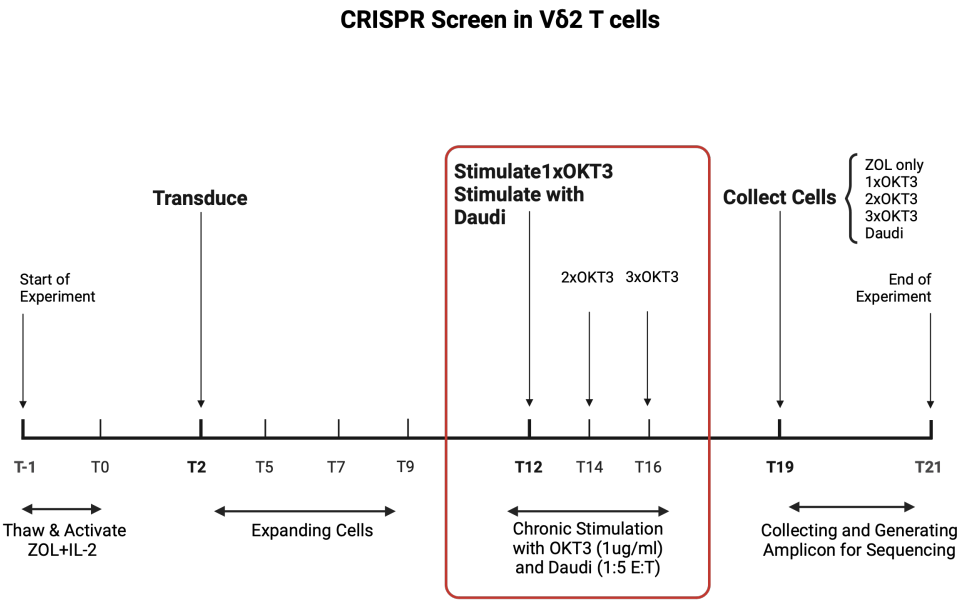



*Figure 2.3: An experimental time-line for CRISPR screen in Vδ2 T cells.*

## Transduction of Vδ2 T cells

Before transduction, frozen Vδ2 T cells were thawed and rested overnight then stimulated with Zol and IL-2 two days before transduction. Multiple protocols were tested for Vδ2 T cells transduction; the most efficient is coating 1ug/cm$^2$ retronectin (Takara Bio) in a non-treated flat bottom 96-well plate (Corning) a day before transduction. The plate was wrapped with parafilm and left at 4C. The retronectin was aspirated the next day, and the plate was ready for transducing.

For virus thawing, vials were transferred from -80 to dry ice, then thawed immediately within 5-10 minutes before use. Virus (100µl) was added to retronectin-coated plates spun in a centrifuge for 2 hours @ 32°C and at 2000 rpm. Following the spin, the supernatant was immediately removed. Cells added at the required number were added before the plates were dried. The plates was then centrifuged again to increase contact of the virus to cells for 10-15min at 400xg at 37°C. The plates were left in the incubator (at 37°C).

GFP is checked two days after transducing the cells, as they should be green. The cell's media was replenished with fresh IL-2 and left for another day in culture. The cells were then split 1:2 and kept in culture for an extra day, where it was replenished with IL-2 and fresh media. After ten days of expanding the cells, transduced Vδ2 T cells were used to set the required assays.

On day 12, some cells were stimulated with the first OKT3 and some with Daudi targets at an effector-to-target ratio of 1:5, followed by a second OKT3 addition on day 14 to the OKT3 stimulated samples and final stimulation of OKT3 on Day 16 to the two times OKT3 stimulated samples. The cells were left in culture for 3 additional days before analysis and collection. The final conditions were Zol only Vδ2 T cells, Zol and OKT3 stimulated Vδ2 T cells, Zol and 2x OKT3 stimulated Vδ2 T cells, Zol and 3x OKT3 stimulated Vδ2 T cells and Zol with Daudi stimulated T cells at 1:5 target-to-effector ratio. On day 19, the cells were collected for analysis using FACS and pelleting for DNA amplicon generation using polymerase chain reaction (PCR). The pellets were stored at -80°C for DNA extraction and PCR for Next-generation sequencing. Full time-line is shown in Figure 2.3.

### DNA Extraction and Amplicon Generation

DNA extraction was performed using the DNeasy Blood & Tissue Kit (Qiagen, Cat. No. 69504), and the protocol was followed as described in the kit's manual.

DNA amplicons were generated using Phusion® High-Fidelity PCR Kit (New England BioLabs, Cat. No. E0553S). The primers used are shown in Table 2.11. PCR reaction set-up and thermos cycler program are shown in Table 2.12 and Table 2.13, respectively.

*Table 2.11: Primers used for DNA amplicon generation.*

| Primer | Sequence | Length (nucleotides) | GC content | Tm | Annealing Tm |
|---|---|---|---|---|---|
| **Forward** | 5` TTA GGC AGG GAT ATT CAC CAT 3` | 21 | 42.9% | 53.3°C | 62°C |
| **Reverse** | 5` TAC TAT TCT TTC CCC TGC AC 3` | 20 | 45% | 51.7°C | 62°C |

*Table 2.12: PCR reaction set-up*

| COMPONENT | 50 µl REACTION | FINAL CONCENTRATION |
|---|---|---|
| **Phusion DNA Polymerase** | 0.5 µl | 1.0 units |
| **5X Phusion HF** | 10 µl | 1X |
| **10 µM Forward Primer** | 2.5 µl | 0.5 µM |
| **10 µM Reverse Primer** | 2.5 µl | 0.5 µM |
| **10 mM dNTPs** | 1 µl | 200 µM |
| **Template DNA** | variable | < 250 ng |
| **Nuclease-free water** | to 50 µl | |

*Table 2.13: Thermocycler Conditions*

| CYCLE STEP | CYCLES | TEMP | TIME |
|---|---|---|---|
| **Initial denaturation** | 1 | 98°C | 30 seconds |
| **Denaturation**<br>**Annealing**<br>**Extension** | 30 | 98°C<br>62°C<br>72°C | 5-10 seconds<br>10-30 seconds<br>15-30 seconds/kb |
| **Final extension** | 1 | 72°C | 5–10 minutes |
| **Hold** | 1 | 4°C | ∞ |

## Analysis of sequencing results and target hits identification

CRISPR screen validation was sent to Genewiz Azenta for unique sequence identification and relative abundance calculations from amplicon sequencing using Illumina MiSeq 2x250 bp sequencing with 10-15 million raw paired-end reads per flow-cell and Phi-X Spike-in. Prepared PCR reactions of the CRISPR library were sent at a concentration of 20ng/µl with a minimum of 500ng per sample.

# Cell Gating Strategy

### *Cell Gating Strategy for Chronic Stimulation Models*

The gating strategy for T cells following FACS for αβT cells and Vδ2 T cells was as follows. All lymphocytes were initially gated, followed by live cells and then single cells. Finally, all αβ TCR-positive and CD3-positive cells were gated as αβ T cell population. All Vδ2 TCR-positive cells and CD3-positive cells were gated as Vδ2 T cells. The required analysis is performed on each of the gated populations (Figure 2.4).

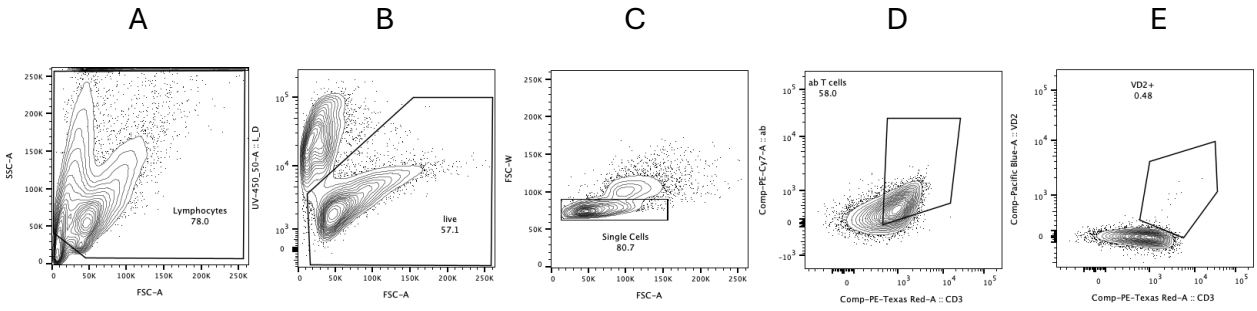



*Figure 2.4: Cell gating strategy for αβ and Vδ2 T cells following selection of live lymphocytes and single cells. (A) shows lymphocytes gate with FSC-A (forward scatter – area on the x-axis) and SSC-A (side scatter-area on the y-axis).(B) shows selection of live cells using viability dye (L_D on the y-axis) and FSC-A (on the x-axis).(C) single cells are selected from FSC-A (on the x-axis) and FSC-W (Forward scatter-width on the y-axis).(D)shows αβ T cells gate by selecting CD3-positive (CD3 on the x-axis) and ab TCR positive cells (ab in the y-axis). (E) Vδ2 T cells are selected with CD3-positive (on the x-axis) and Vδ2 TCR positive cells (VD2 on the y-axis).*

### *Cell Gating Strategy for FACS sort of Vδ2 cells prior to bulk- RNA sequencing*

The gating strategy for sorting Vδ2 T cells on day 6 of expansion was as follows. Following the gating of all Live cells and single cells, Vδ2 positive cells were selected to be sorted as a pure population for bulk-RNA sequencing (Figure 2.5).

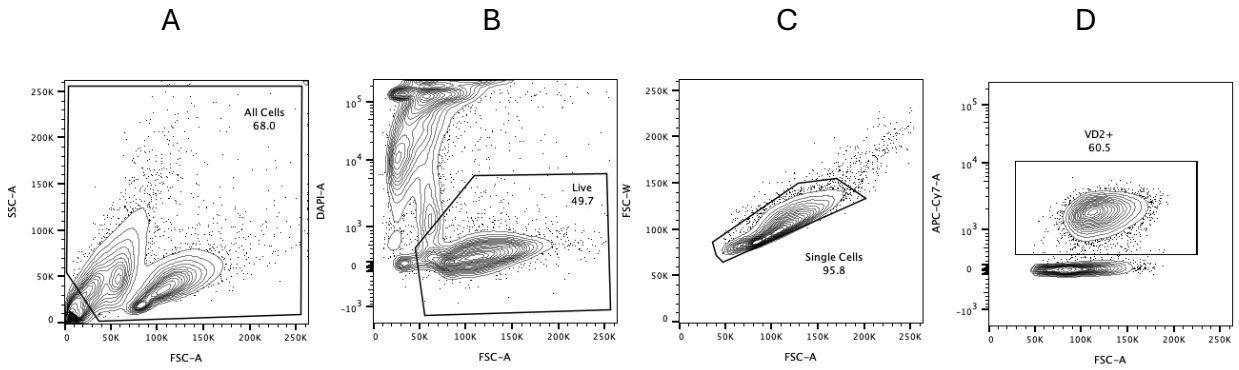



*Figure 2.5: Cell gating strategy for sorting Vδ2 T cells prior to RNA sequencing. (A) shows gating of all cells with FSC-A (forward scatter–area on the x-axis) and SSC-A (side scatter-area on the y-axis).(B) shows selection of live cells using viability dye (DAPI on the y-axis) and FSC-A (on the x-axis).(C) single cells are selected from FSC-A (on the x-axis) and FSC-W (Forward scatter-width on the y-axis).(D) shows selection of all Vδ2 TCR positive cells (APC-Cy7 positive cells on the y-axis) and using FSC-A (on the x-axis)).*

### *Cell gating strategy for Vδ2 T cells for functional assays and CRISPR screen*

Cells are initially gated for all lymphocytes, followed by live cells and then single cells. Finally, all Vδ2-positive cells are gated and followed by any phenotyping required (Figure 2.6).

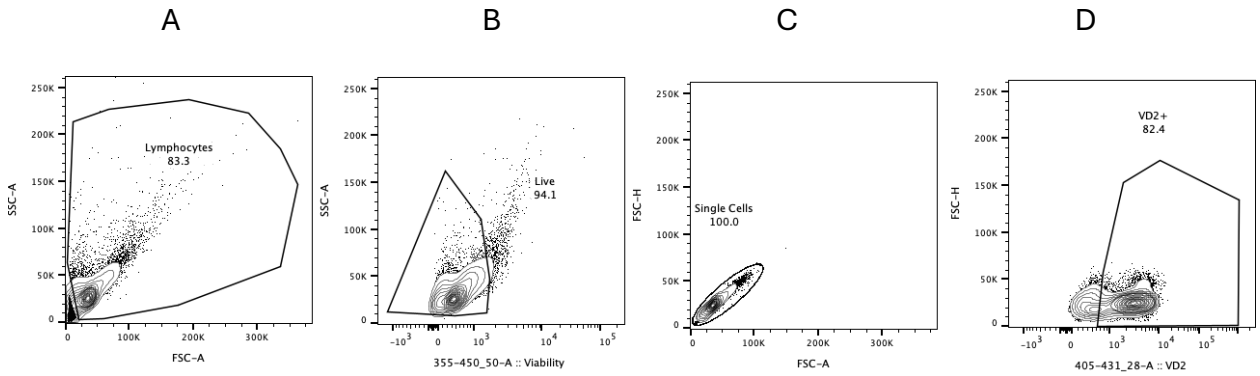



*Figure 2.6: Cell gating strategy for Vδ2 T cells used in functional assays and crispr screen. (A) shows gating of lymphocytes with FSC-A (forward scatter–area on the x-axis) and SSC-A (side scatter-area on the y-axis).(B) shows selection of live cells using viability dye (viability on the x-axis) and SSC-A (on the x-axis).(C) single cells are selected from FSC-A (on the x-axis) and FSC-H (Forward scatter-height on the y-axis).(D) shows selection of all Vδ2 TCR positive cells (VD2 on the x-axis) and using FSC-H (on the y-axis)).*

# Chapter 3

## Chronic Stimulation models to study exhaustion in Vδ2 T cells

### Introduction and Summary

T cell exhaustion has been identified as a major challenge for T cell therapeutics. In conventional T cells (αβ T cells), the study of exhaustion has led to insights into its aetiology, and potential therapeutic approaches to overcome it.

Gamma delta (γδ) T cells are a subtype that is therapeutically promising due to their unique characteristics and ability to bridge innate and adaptive immunity. However, compared to αβ T cells, T cell Exhaustion is not well understood in γδ T cells, of which the Vγ9Vδ2 subtype is the most prevalent. There is a need to establish an understanding of how exhaustion can affect the behaviour and functionality of these cells.

We developed two *in vitro* chronic stimulation models using combinations of Zoledronic acid (ZOL) and OKT3 anti-CD3 monoclonal antibody, both of which stimulate via the Vγ9Vδ2 T cell receptor. Model 1 involves three sequential stimulations using OKT3 (on day 0 with/without days 3 and/or 6). Model 2 involves an initial stimulus-specific for the Vγ9Vδ2 TCR (ZOL) with or without two subsequent OKT3 stimulations (on days 3 and/or 6). A third model (model-3) using Daudi target cells was used to stimulate Vδ2 T cells using three stimulations sequentially (on day 0 with/without days 3 and/or 6).

Model-1 and model-3 did not induce differences in absolute cell numbers or expression of exhaustion markers between the conditions tested. However, model-2 induced significantly lower absolute T-cell numbers in conditions with multiple stimulations (double or triple) than single stimulated conditions. Moreover, significant upregulation of exhaustion markers (TIM-3, LAG-3, TOX and TCF-1) was noticed with multiple stimulated conditions compared to single stimulation. A proliferation study identified the emergence of heterogeneous populations on day 6 (following double stimulations) including hypo-proliferating putative exhausted Vγ9Vδ2 T cells.

## Developed *in vitro* Stimulation Models

Our main interest is to study whether human γδ T cells develop exhaustion following chronic stimulation as seen in cancer. In vivo mouse models using human γδ T cells are difficult to produce, thus, an in vitro chronic stimulation model is required to assess exhaustion in human γδ T cells. The standard model used to study exhaustion in the context of chronic infection is the Lymphocytic choriomeningitis (LCMV) in vivo model. A thorough optimization of a suitable *in vitro* infection model will be required to evaluate exhaustion in Vδ2 T cells. Also, there is a need to choose optimal time points for phenotyping the T cells for exhaustion and exhaustion-associated markers. Recent publications will impact the model design on exhaustion phenotyping in T cells. Moreover, this work aims to study γδ T cells which have not previously been studied in the context of T cell exhaustion.

The aim to develop an efficient model for exhaustion characterization was initiated by choosing a suitable stimulus. Two models of chronic stimulation have been evaluated. In the first model, Zoledronic acid (ZOL) was used to initiate the stimulation of Vγ9Vδ2 T cells and followed by repeated (one or two) stimulations using CD3 cross-linking with the OKT3 antibody. The ZOL-based model provides a more transient activation of Vγ9Vδ2 T cells and stimulates them via monocyte-dependent degranulation. In detail, ZOL induces degranulation of monocytes in peripheral blood mononuclear cells, following farnesyl diphosphate synthase (FPPS) inhibition in the mevalonate pathway. FPPS inhibition leads to the accumulation of IPP (phospho-antigens) and triggers a conformational change of the butrophyllin receptor domains (BTN3A1 and BTN2A1) on monocytes, leading to the recognition by Vγ9Vδ2 T cells (Schematic 3.1 (B)) (Karunakaran et al., 2020). Hence, ZOL activation is transient and stops immediately after the killing of monocytes in cultures.

The second model was developed to give single, double or triple stimulations of OKT3 to stimulate both Vγ9Vδ2 and αβ T cells (Schematic 3.1 (A)). OKT3 is a monoclonal antibody used for T cell activation. OKT3 is very potent and can substantially activate T cells at

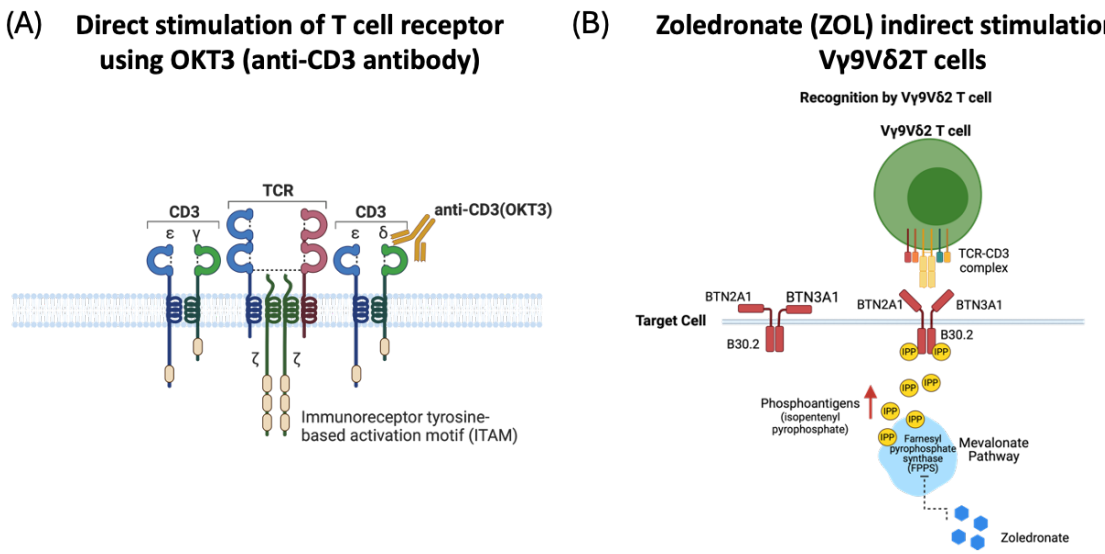
concentrations as low as 10ng/ml. In contrast to ZOL, OKT3 is very stable in the media and persists for more prolonged durations after stimulation. Moreover, the use of OKT3 as a stimulus will allow the comparison of Vγ9Vδ2 to αβ T cells as it can stimulate all types of T cells (Van Wauwe, J. P., De Mey, J. R., & Goossens, J. G, 1980).

The third model, Daudi cells were used to stimulate Vγ9Vδ2 T cells three sequential stimulations on day 0 with/without days 3 and/or 6. Daudi cells are B cell lymphoma and are highly sensitive well-known as efficient stimulators of Vγ9Vδ2 T cells due to their high level accumulation of IPP (Cano et al., 2021).

## Experimental limitations

In developing the chronic stimulation models, we used concentrations of stimulants (OKT3 and Zoledronate), which were widely reported as effective in stimulating T cells preclinically and in the clinic (Boucher et al., 2023; Glinos et al., 2020). Standard concentrations are required when comparing the literature and recent reports to interpret the results better.

Testing the use of variations of concentrations could provide better insights into the responses of T cells following expansion. Moreover, small pilot studies could incorporate looking at different intervals between stimulations while using a range of stimulant concentrations to provide more knowledge about T cell expansion and activation. However, this was not possible when the model was developed due to the resource-intensive nature of evaluating the different models with the number of biological replicates and cells available.




*Schematic 3.1: T cell stimulation using OKT3 (anti-CD3) is shown in (A) and Vγ9Vδ2 T cells indirect stimulation using Zoledronate via monocytes dependent degranulation is shown in (B).*

## OKT3 based chronic Co-stimulation of Vγ9Vδ2 and αβ T cells

### *Time-course of in vitro chronic stimulation model*

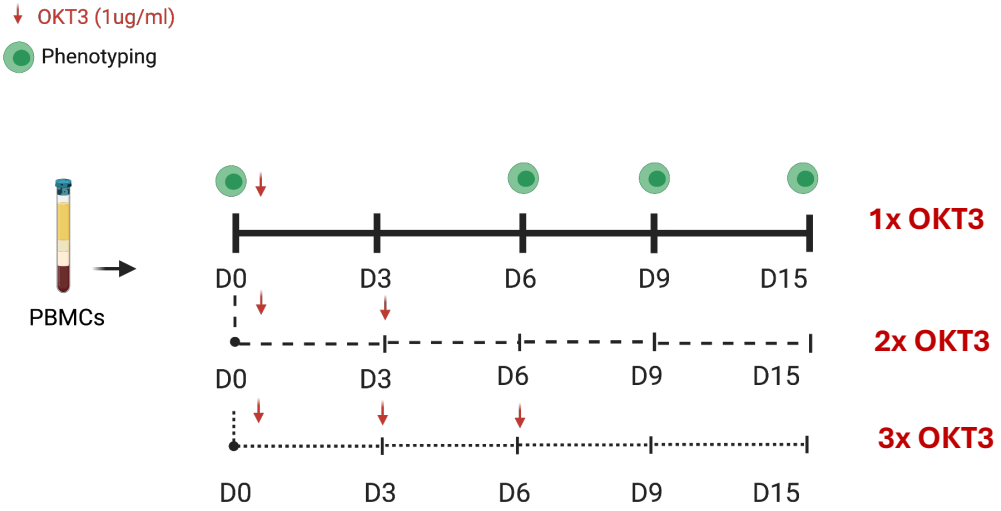



*Figure 3.1: Schematic representation of Vγ9Vδ2 T cells and αβ T cells OKT3 induced chronic stimulation model in vitro. The stimulus used is OKT3 (1ug/ml) at Day 0 after the isolation of PBMCs from blood buffy coat for the first condition and followed by additional OKT3 (1ug/ml) stimulation at day 3 for a second condition or days 3 and 6 for a third condition. Samples with single, double and triple OKT3 stimulations are collected for phenotypic analysis on Day 0 (baseline), Day 6, Day 9 and Day 15. Three biological replicates were used.*

To better understand the success of the chronic model, it is essential to compare the results with recent publications. However, there are minimal studies with Vγ9Vδ2 T cells and abundant work on αβ T cells. For that reason, we included the phenotyping of αβ T cells in this chronic stimulation model as there are many publications on the different phenotypes of exhausted αβ T cells-(Khan et al., 2019). OKT3 is a well-studied activator of T cells, and therefore it was used to stimulate all T cells in cultures (αβ and Vγ9Vδ2 T cells). The timing for phenotypical analysis was developed to look at the phenotypical changes occurring to cells up to day 15 after stimulations on days 0, 3 and 6 (Figure 3.1). Moreover, additional questions regarding cell death during the induced time-course stimulation were investigated.

The first dose of stimulus is given at day 0 using OKT3 (1µg/ml). Then, stimulations of OKT3 (1µg/ml) are provided on days 3 and 6, resulting in 3 different conditions (i.e., single, double and

triple stimulated cells). The cells are then characterised phenotypically at specific time points and analysed for their specific characteristics. The time course includes phenotypic analyses using Fluorescence-activated cell sorting (FACS) at day 0 (before stimulation), day 6 (post 1st and 2nd OKT3 stimulation), day 9 (post 1st, 2nd and 3rd OKT3 dose), and day 15 (post 1st, 2nd and 3rd treatment with OKT3). Cells are counted at each time point to assess the impact of chronic stimulation on cell numbers in cultures and after each stimulation addition.

*Effect of chronic stimulation model on cell numbers and TCR expression*

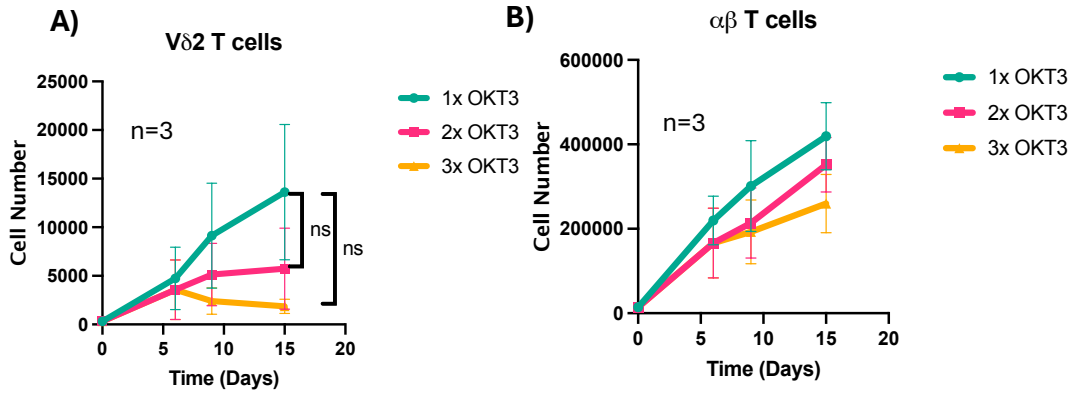



*Figure 3.2: Cell numbers drop with increasing stimulation during the OKT3 induced chronic stimulation. PBMCs cultures were stimulated with OKT3 (1ug/ml) at day 0 and sub proportions were also stimulated with a second OKT3 (1ug/ml) dose at Day 3 or with a second and third doses at Day 3 and Day 6. Comparisons between cultures only stimulated with single OKT3 (1x), at Days 0,6,9 and 15, double OKT3 (2x) , at days 6, 9 and 15, and triple OKT3 (3x), at Day 6, s 9 and 15, were made in regards of Total numbers and are shown for A) Vγ9Vδ2 T cells and B) αβT cells at days 0,3,6 and 9. Statistics analysis was performed using two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

In this study, we first looked at changes in cell numbers during the time-course (Figure 3.2). For all conditions except the triple stimulated Vγ9Vδ2 T cells, the number of cells increases with time for Vγ9Vδ2 T cells and αβ T cells. However, with increasing stimulations on days 9 and 15, the cell numbers relatively decrease as the stimulation increases. Proportions of Vγ9Vδ2 T cells in the blood are highly variable and are much lower than αβ T cells. As expected, the number of αβ T cells surpass Vγ9Vδ2 T cells at all time points due to the over-dominating proportions of αβ T cells from day 0 and in all donors tested. Although the trend shown with differences in cell numbers is similar between the αβ and Vγ9Vδ2 T cells, the difference is more profound with Vγ9Vδ2 T cells.

The difference in cell numbers between single and multiple stimulations suggested this model might be successful in inducing a functional hyporesponsiveness secondary to excessive activation. However, the decrease in expanded T cell numbers following increased stimulation does not reach statistical significance. To better assess the success of this model, phenotypical analysis of cultures is required at the specific times shown.

Bearing in mind that cell counting was based on cells positive for TCR and CD3 staining, gating of T cells is affected by TCR negative or TCR dim populations, which might be caused by TCR internalisation following stimulation. Reduced cell numbers might be either due to an artefact caused by undercounting cells as a result of TCR internalization from persistent stimulation or activation-induced cell death in stimulated cultures.

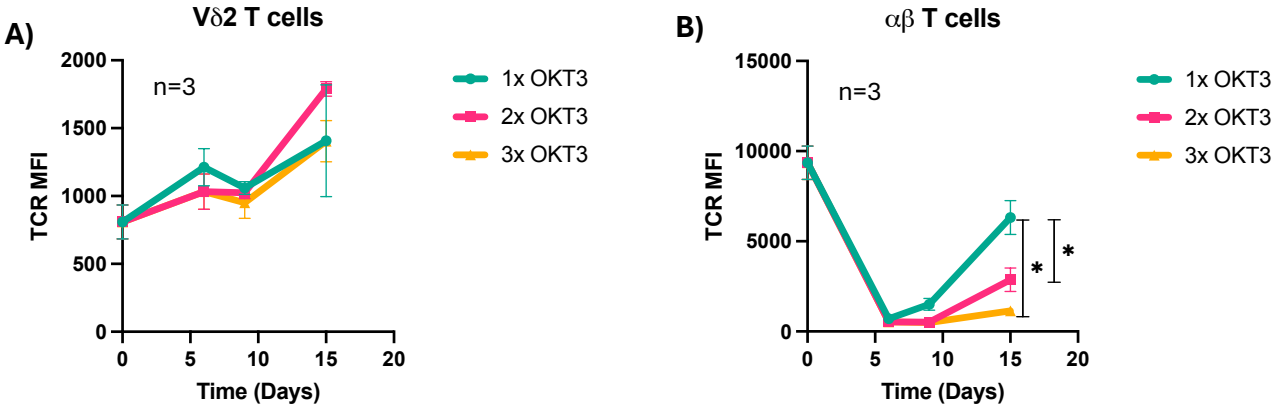



*Figure 3.3: Trends of the T cell receptor (TCR) surface expression levels of Vγ9Vδ2 T cells and αβT cells with increasing stimulation during an* OKT3 induced chronic stimulation model. *PBMCs cultures were stimulated with OKT3 (1ug/ml) at day 0 and sub proportions were also stimulated with a second OKT3 (1ug/ml) dose at Day 3 or with a second and third doses at Day 3 and Day 6. Comparisons between cultures only stimulated with single OKT3 (1x), at Days 0,6,9 and 15, double OKT3 (2x) , at days 6, 9 and 15, and triple OKT3 (3x), at Days 6, 9 and 15. Expression levels were plotted using the mean fluorescence intensities (MFI) of TCR expression on the surface of A) Vγ9Vδ2 T cells and B) αβT cells in cultures. Statistics analysis was performed using two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

Using the OKT3 model, we evaluated TCR expression changes in Vγ9Vδ2 T cells and αβ T cells and compared the changes in their patterns. Looking at the levels of TCR expressed on the surface, overall decrease of TCR levels were seen on Day 6 for all conditions, and we noticed a very significant decrease in TCR levels on αβ T cells on day 15 and specifically when comparing single stimulation to multiple stimulation conditions while slight decrease was observed on Day 9 but not Day 6 with the multiple stimulated conditions (Figure 3.3). There is a

pronounced decrease in TCR levels between single-stimulated and double-stimulated cultures and single and triple-stimulated samples. However, there were no significant differences in TCR expression for Vγ9Vδ2 T cells between single and multiple stimulated conditions. In general, TCR levels in Vγ9Vδ2 T cells increased on day six and continued to increase between day 9 and day 15 which is a contrast of the decrease in TCR expression observed with αβ T cells. The change in TCR expression of αβ T cells was statistically significant on Day 15 between the single and multiple stimulated samples, but this was not the case for Vγ9Vδ2 T cells.

The experiment revealed unexpected findings regarding the TCR responses of αβT and γδ T cells, which, when subjected to OKT3 stimulation, exhibited remarkably different responses in terms of the MFI of the T cell receptor.  OKT3 is an antibody that binds the epsilon subunit of CD3. It is well known that OKT3 engagement can lead to downregulation of αβTCR expression following the addition of the antibody, as was observed in our studies (Fig 3.3b). In cells where only a single OKT3 is used for stimulation, the MFI increases again following transient decrease. However, as expected, when two or three OKT3 stimulations were used, the cells' MFI did not increase.

In contrast, γδ TCR expression was found not to downregulate following OKT3 stimulation, which is surprising since the antibody successfully stimulated the cells. It is known that the configuration of CD3 components is different in gamma delta cells compared with αβ T cells, and this may lead to alternate cross-linking of the TCR such that it signals but does not internalise (Shah et al., 2021). To our knowledge, no previously reported studies describe the absence of γδ TCR internalisation following OKT3 binding.

*Effect of chronic stimulation model on the expression of exhaustion markers*

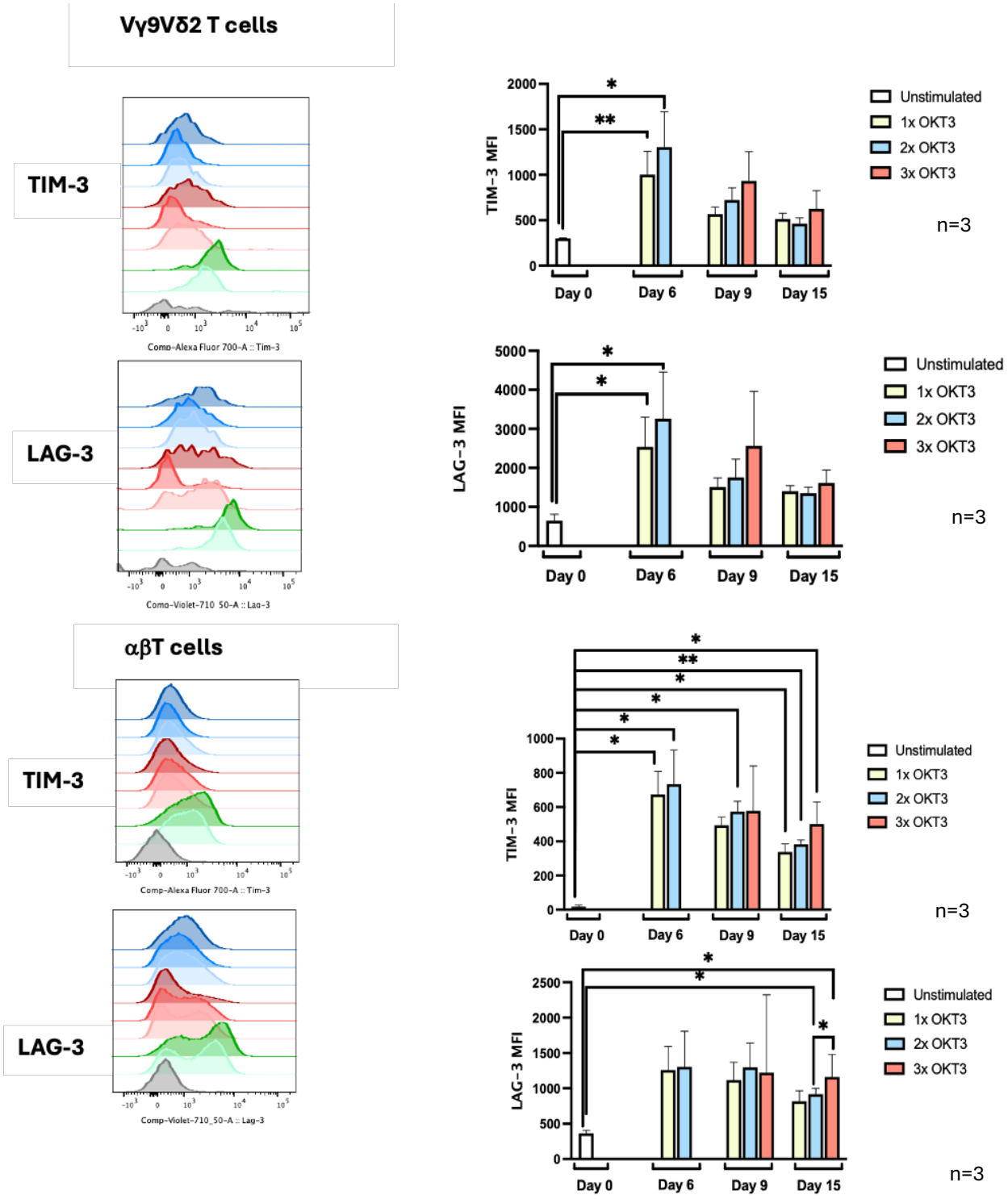



*Figure 3.4: Differential expression of Classical exhaustion markersTIM-3 and LAG-3 in A) Vγ9Vδ2 T cells and B) αβ T cells during a time-course OKT3 induced chronic stimulation model. Expression levels (mean fluorescence intensities) on cultures pre-stimulation and on days 9 and 15 are shown on the right. Histograms of cell counts are shown on the left. Statistics analysis was performed using two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

Levels of TIM-3 and LAG-3 were measured during the time-course following OKT3 stimulation model *in vitro* (Figure 3.4). For both Vγ9Vδ2 T cells and αβ T cells, exhaustion markers were significantly upregulated at all time points compared to the baseline (pre-stimulation) condition. Also, with both Vγ9Vδ2 T cells and αβ T cells, there is a similar trend for all markers (TIM-3 and LAG-3), where the upregulation levels correlate with the number of stimulations provided at all 3-time points (days 6,9 and 15). However, the difference between the different stimulation conditions did not reach statistical significance in most comparisons. The only significant difference was LAG-3 expression in αβ T cells between double and triple-stimulated conditions. The upregulation of TIM-3 and LAG-3 is consistent with their roles as activation markers and checkpoints to avoid excessive response. The lack of significant differences between single and multiple stimulations indicates the failure of this model to provide evidence of induced exhaustion following repeated stimulation in Vγ9Vδ2 T cells and αβ T cells.

### *Effect of chronic stimulation model on cell death*

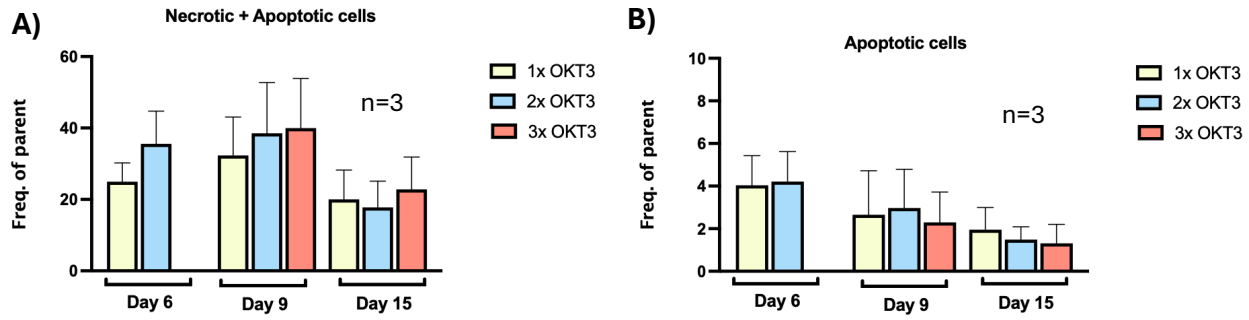



*Figure 3.5:* Effect of chronic stimulation model on cell death. Percentages of A) total of apoptotic and necrotic and B) apoptotic *Vγ9Vδ2* T cells during OKT3 induced chronic stimulation *in vitro*. *Data presented for unstimulated cultures, cultures only stimulated with single OKT3 (1x) (on days 6, 9 and 15), double stimulated cultures with OKT3 (2x) (on days 9 and 15), and triple stimulated cultures with OKT3 (3x) (on Days 6, 9 and 15). Three biological replicates were used.*

The reduction in cell numbers with multiple stimulations conditions was interesting and we wanted to understand whether the effect of the reduction was associated with increased cell death and is caused by the phenomenon of activation-induced cell death. To gain data on potential induced cell death, the cultured cells were analyzed for their necrotic and apoptotic characteristics based on gating PI-positive cells as necrotic and annexin-V positive and PI-

negative cells as apoptotic (Figure 3.5). Necrosis and apoptosis are both death mechanisms. Necrosis is a pathological cell death process that appears due to changes in the environment or the presence of inflammation. Necrotic cell death could elicit inflammatory responses.

Apoptosis is a programmed cell death that is autonomous to the cells. Apoptosis is a physiological process and is not associated with the release of inflammatory molecules. The results show that very few cells were identified as apoptotic at all time points. The only explanation is that most cells in the studied populations were necrotic. However, there is no significant difference between the number of actively dying cells at any time. Therefore, there is no evidence suggesting a difference in percentages of dead cells to explain any differences in total cell numbers between single and multiple stimulations.

In summary, the triple OKT3 model did not induce any differences in the expression of exhaustion markers (TIM-3 and LAG-3) in Vγ9Vδ2 T cells. Moreover, the differences in cell numbers between the conditions tested did not reach statistical significance. The triple OKT3 stimulation model was not strongly suggesting the presence of an exhaustion-like phenomenon in Vγ9Vδ2 T cells. Therefore, alternative exhaustion-inducing stimulation models will be explored.

## Zoledronic acid and OKT3-based chronic stimulation of Vγ9Vδ2 T cells

*Time-course of ZOL/OKT3 in vitro chronic stimulation model*

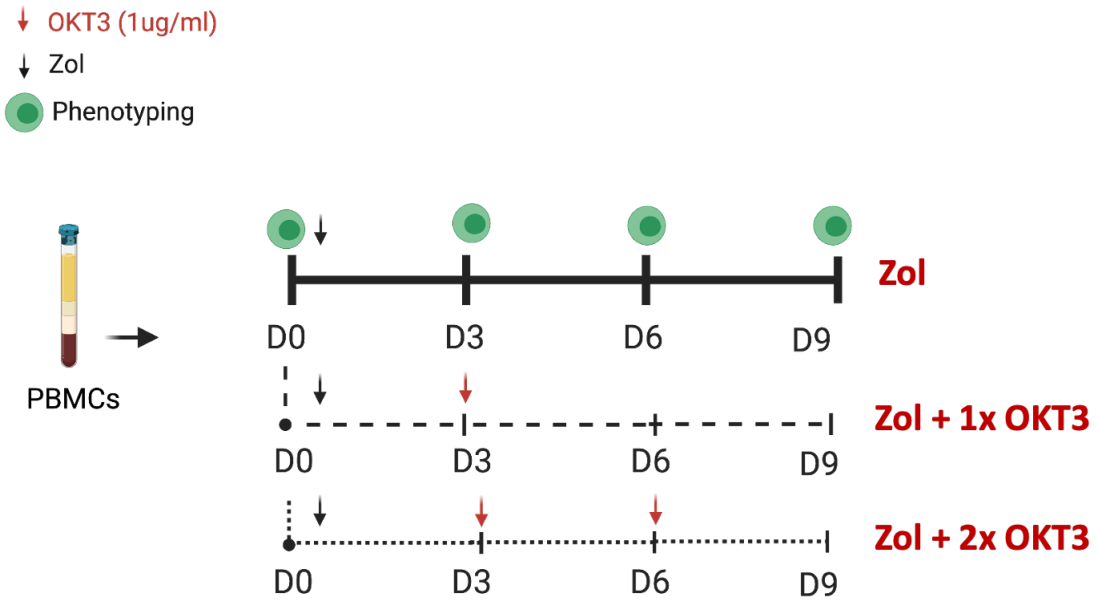



*Figure 3.6: Schematic representation of Vγ9Vδ2 T cells chronic stimulation model in vitro. The stimulus used is Zoledronic acid (ZOL) on Day 0 after the isolation of PBMCs from blood buffy coat and followed by an OKT3 (1ug/ml) stimulation on days 3 and 6. Samples with ZOL only stimulation, ZOL and single OKT3 stimulation and ZOL with double OKT3 stimulations are collected for phenotypic analysis on Day 0, Day 3, Day 6 and Day 9.*

Next, we investigated an alternative stimulation model using an initial weaker stimulus followed by OKT3 stimulations. We wanted to test whether using a weaker stimulus initially would lead to a more profound difference in cell numbers between single and multiple stimulations. The chronic stimulation was induced *in vitro* using Zoledronic acid and OKT3 (Figure 3.6). The stimulation model was specifically designed to trigger chronic activation of Vγ9Vδ2 T cells and investigate the immunophenotypes rising from this setting. The stimulation was done post the isolation of Peripheral blood mononuclear cells (PBMCs) from a blood buffy coat. The first stimulus is given at day 0 using Zoledronic acid at a final concentration of 5μM with IL-2 (100 IU/ml). Then, OKT3 (1 μg/ml) was provided to stimulate cultures on days 3 and 6. IL-2 was added to the cells every 2-3 days. The cells phenotypes are then investigated at specific time

points and characterized for their specific features. The time-course study includes a phenotypic analysis by FACS at day 0 (pre-stimulation), day 3 (post-ZOL stimulation and before first OKT3 dose), day 6 (post-first OKT3 dose and before 2nd OKT3 dose) and day 9 (post-ZOL treatment and 2 OKT3 doses). The cells are also counted at each time point to assess the impact of chronic stimulation on cell numbers in cultures.

### *Effect of ZOL/OKT3 chronic stimulation model on cell numbers and TCR internalization*

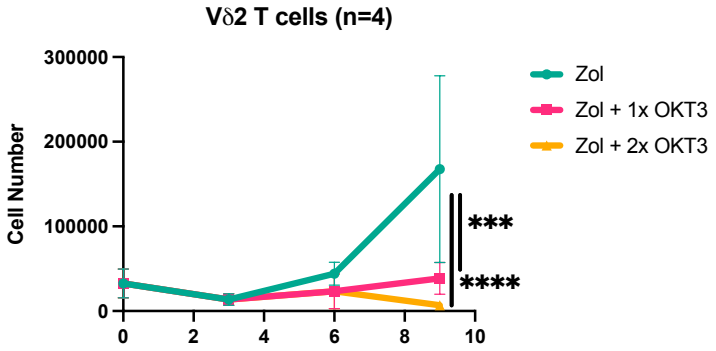



*Figure 3.7: Cell numbers drop with increasing stimulation during the ZOL/OKT3 induced chronic stimulation. PBMCs cultures were stimulated with ZOL (5uM) at day 0 and sub proportions were also stimulated with OKT3 (1ug/ml) dose at Day 3 or with a second dose at Day 6. Comparisons between cultures only stimulated with single stimulus (ZOL), double stimulus (ZOL and 1x OKT3), and triple stimulus (ZOL and 2x OKT3) were made at Days 0,3,6 and 9 in regards of Total numbers and are shown for Vγ9Vδ2 T cells. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Four biological replicates were used.*

The absolute cell numbers were counted to measure the impact of chronic activation on T-cell expansion (Figure 3.7). In the studied cultures, cell numbers were shown to drop slightly on day 3, then increase on day 6 (post-ZOL stimulation on day 0) for the ZOL-only stimulated condition. Cell numbers of ZOL with single OKT3 and ZOL with double OKT3 stimulation conditions drop on day 3 and increase on day 6. However, on day 9, the cell count was lower in the condition with multiple stimulations.

The 10-fold difference in mean cell number between single and multiple stimulations is consistent with successfully inducing functional hyporesponsiveness following the cells' chronic activation.

The chronic model must be further validated for its role in inducing chronic activation in T cells by the phenotypic analysis of cells during the time course of the study. Moreover, as described before, cell numbers could be reduced due to miscalculation of cell counts caused by TCR internalisation, as TCR and CD3 staining are used in identifying the cells. TCR internalisation is an expected response from T cells exposed to chronic activation. An alternate explanation for the decrease in cell numbers could be activation-induced cell death, another potential phenomenon caused by the activation of T cells, or exhaustion.

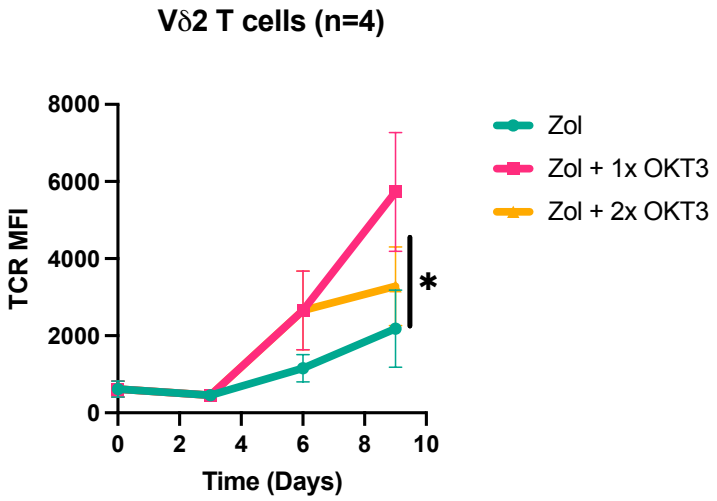



*Figure 3.8: Trends of the T cell receptor (TCR) surface expression levels of Vγ9Vδ2 T cells with increasing stimulation during ZOL/* OKT3 induced chronic stimulation model. *PBMCs cultures were stimulated with ZOL (5uM) at day 0 and sub proportions were also stimulated with OKT3 (1ug/ml) dose at Day 3 or with a second dose at Day 6. Comparisons between cultures only stimulated with single stimulus (ZOL), double stimulus (ZOL and 1x OKT3), and triple stimulus ZOL and 2x OKT3) were made at at Days 0,3,6 and 9. Expression levels were plotted using the mean fluorescence intensities (MFI) of TCR expression on the surface of Vγ9Vδ2 T cells. Statistical analysis was done using a two-way ANOVA and Significance shown for P values <0.05. Four biological replicates were used.*

We looked at differences in TCR expression to validate that TCR internalisation does not affect our results for differences in cell numbers. Downregulation of TCR is common in some models and it is crucial to ensure that the cell counts are not under-represented due to TCR internalisation.

The analysis (Figure 3.8) showed that the levels of TCR expressed on the surface of Vγ9Vδ2 T cells are not decreasing with increased stimulation, similar with what is shown before with the

OKT3 model for Vγ9Vδ2 T cells (Figure 3.3A) and confirms that Vγ9Vδ2 T cells respond differently to ab T cells and stabilizes its TCR expression with increased stimulations. Also, no reduction in TCR levels means that there is no likelihood of under-reporting the numbers of Vγ9Vδ2 T cells and confirms the differences observed with cell counts, validating the superiority of the ZOL model compared to the OKT3 model in reducing cell numbers in multiple stimulated conditions.

### *Effect of ZOL/OKT3 chronic stimulation model on the expression of exhaustion markers*

Several proteins have been identified as markers associated with the development of exhaustion in T cells. These markers include CTLA-4, TIM-3, LAG-3, PD-1 and TIGIT. For this study, as TIM-3 has been shown to be associated with terminal exhaustion in αβ T cells and LAG-3 is also known as an activation and exhaustion marker for T cells, they were chosen to study the phenotypes of cells in the proposed time course *in vitro* using ZOL/OKT3 chronic stimulation model (Figure 3.9). PD-1 was excluded from the study is it didn't show any positive staining with the cells raising the possibility of a technical failure of detection and reducing confidence that the staining antibody was working properly.

There was an upregulation of exhaustion markers (TIM-3 and LAG-3) with multiple stimulated conditions from the baseline (pre-stimulation). A pattern of more significant upregulation of exhaustion markers correlating to the number of stimulations is observed at two-time points (days 6 and 9). A statistically significant difference is noticeable on day 6 (between single and double stimulated conditions) and day 9 (between single and triple stimulated conditions). In contrast to the OKT3 model, the difference in expression was better observed and statistically of more significance.

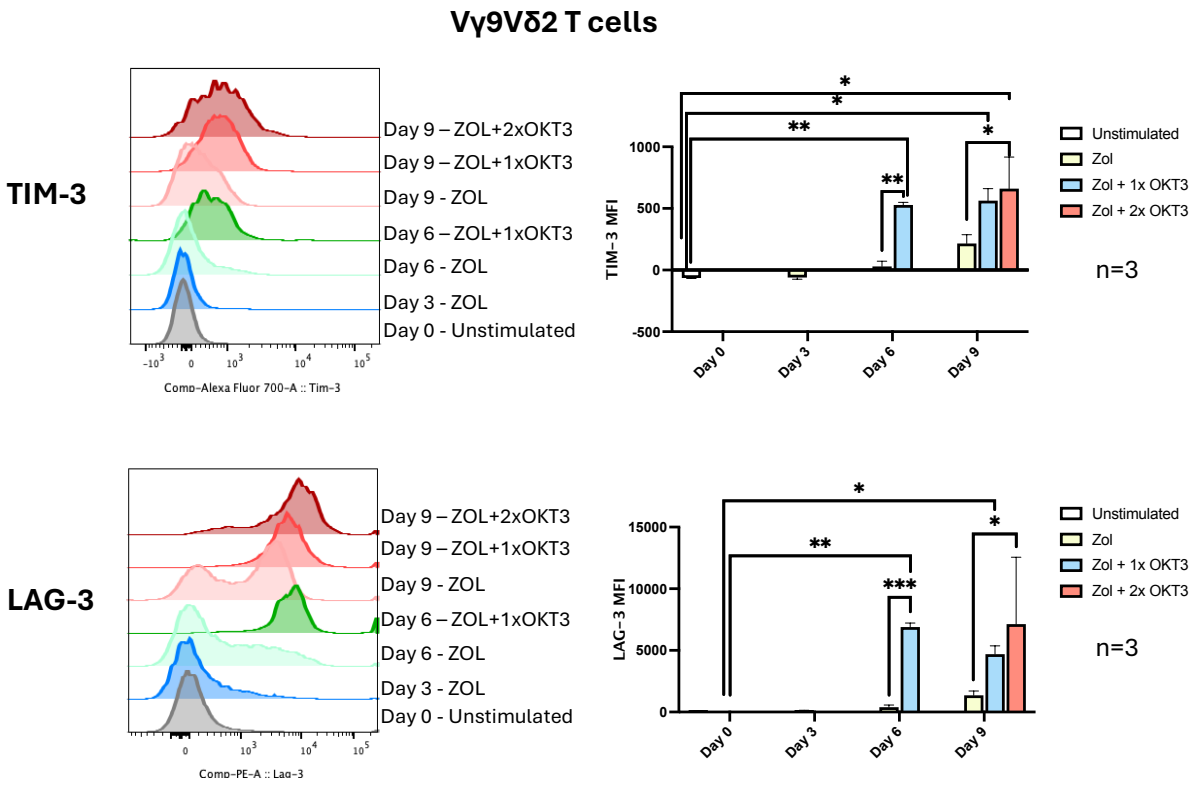



*Figure 3.9: Differential expression of Classical exhaustion markers TIM-3 and LAG-3 on Vγ9Vδ2 T cells during a time-course chronic stimulation model. Expression levels (mean fluorescence intensities) on days 0,3,6 and 9 are shown on the right and histograms of cell counts are shown on the left. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

## *Effect of ZOL/OKT3 chronic stimulation model on the expression of novel exhaustion markers*

Recently, discoveries were made regarding factors associated with T cell exhaustion. Many distinct transcription factors were classified as primary identifiers for exhaustion phenotypes in T cells. One of the key regulators of exhaustion in T cells, which is considered a vital exhaustion-related gene, is TOX. Factors that have been identified as anti-exhaustion were also studied. Anti-exhaustion genes help revive exhausted cells and promote their functionality genetically by the reprogramming of exhaustion-associated gene network (i.e., inhibiting their transcription), alternatively, by displaying inhibitory effects through suppressing protein levels and balancing levels of pro-inflammatory cytokines. A recently recognised transcription factor with the

potential to prevent cellular exhaustion that is most identifiable on stem-like progenitor T cells is TCF-1 (Khan et al., 2019; Beltra et al., 2020). Therefore, TOX and TCF-1 were chosen to phenotype Vγ9Vδ2 T cells in the conditions of repeat stimulation using ZOL and OKT3 (Figure 3.10). TOX and TCF-1 were measured using an intercellular staining technique following permeabilization of the cells to quantify levels of TOX and TCF-1.

An interesting initial observation was that Vγ9Vδ2 T cells showed higher TOX expression levels across time, with the highest levels noticed on day 6. The patterns of TOX expression increased with time and are statistically significant between the baseline (pre-stimulation) condition and day 6 for the multiple stimulated condition. Moreover, there is a marked upregulation of both markers on day 6 between the single and double-stimulated conditions. TOX has been previously reported as an important marker for exhausted T cells and exists at higher levels in terminally exhausted cells. Therefore, the high expression of TOX in the cultures studied under ZOL and OKT3 stimulation is consistent with the emergence of an exhausted population by day 6. Further validating the existence of an exhausted population, TOX expression patterns also match TIM-3 and LAG-3 expression patterns in the same cultures.

Interestingly, TCF-1 levels show similar patterns to TOX expression where an increase in TCF-1 levels was noticed with time (on Day 6) for the multiple stimulated condition. However, the fold change of TOX is greater than TCF-1. As TCF-1 was described as a marker for progenitor stem-like exhausted T cells, it suggests that the cultures following second or third stimulus might contain an increase in the proportion of progenitor-exhausted cells (with high TCF-1 levels) with a representation of a population with low levels of TCF-1 that have transformed into exhausted T cells by day 6. Hence, the data are consistent with a lower number of the exhausted population remains in culture on day 6, with a higher number of progenitor cells resulting in the high level of TCF-1 seen.

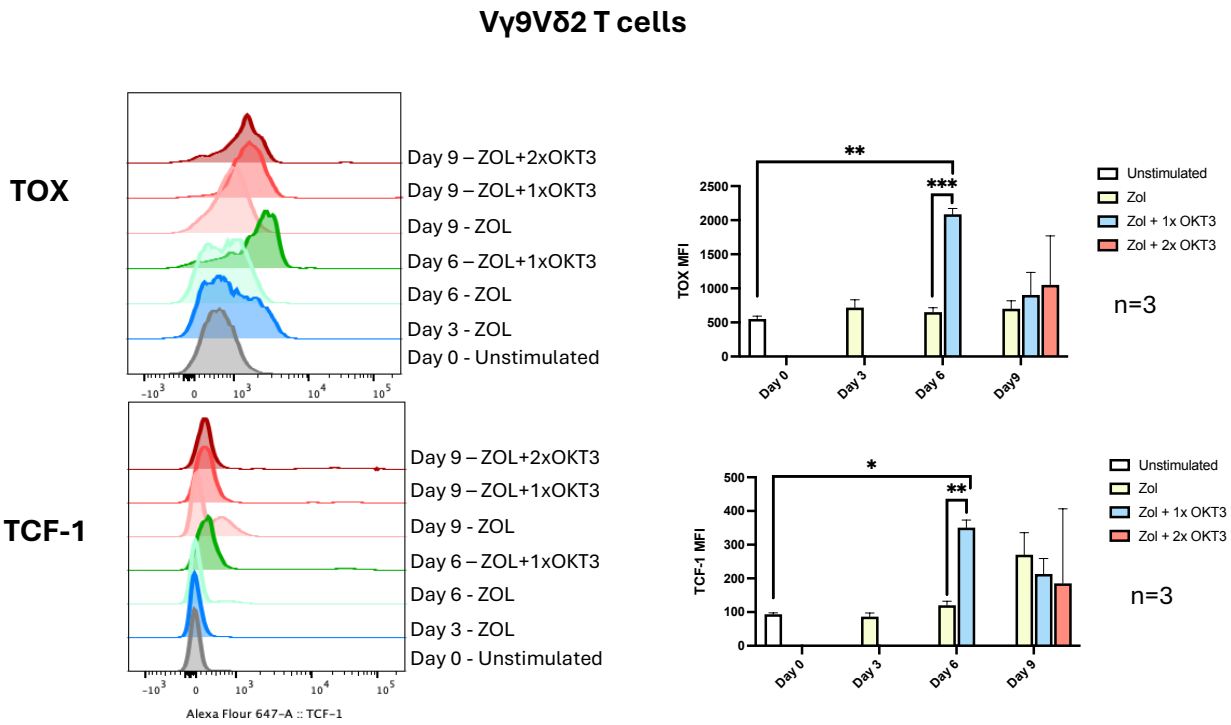



*Figure 3.10: Differential expression of exhaustion markers TOX and TCF-1 on Vγ9Vδ2 T cells during a time-course chronic stimulation model. Expression levels (mean fluorescence intensities) on days 0,3,6 and 9 are shown on the right and histograms of cell counts are shown on the left. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

## *Effect of ZOL/OKT3 chronic stimulation model on cell Proliferation*

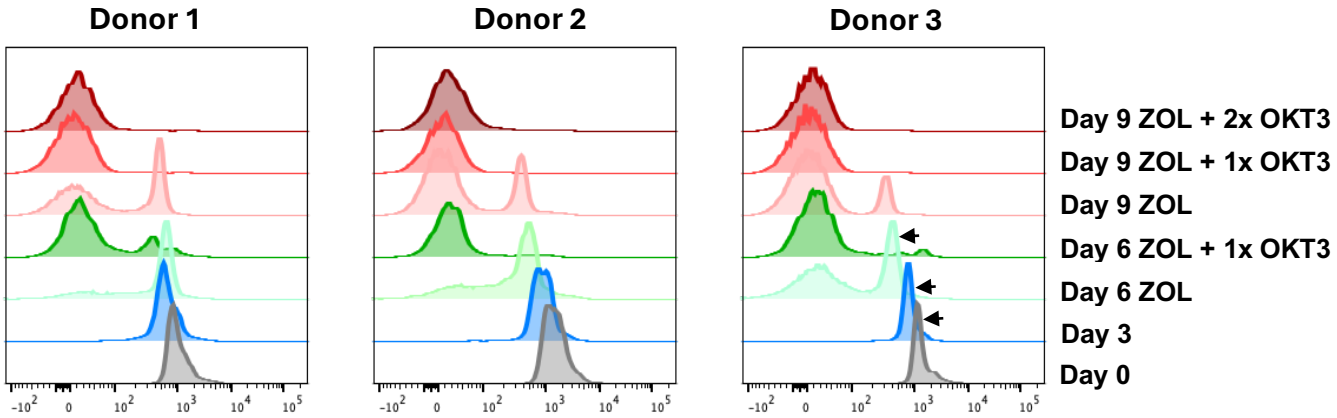



*Figure 3.11: Proliferation capacity of Vγ9Vδ2 T cells during ZOL/*OKT3 induced *chronic stimulation. The figures show proliferation of cells using a proliferation dye (cell trace violet). the assay is analyzed by comparing the peak with the highest intensity (colored with grey illustration) to the diluted peaks with lower intensities (shifted to the left and indicated by an arrow). Each diluted peak represents a single cell division. Multiple peaks in one sample is explained as multiple populations that are presented in the same sample at a single time-point. Three biological replicates were used.*

Additional tests for characterising Vγ9Vδ2 T cells under chronic stimulation were also performed. The proliferation of cells during the time-course study and with different stimulations was studied using a cell trace dilution assay (Figure 3.11). A cell dilution assay measures generations of proliferating cells, the assay is analyzed by comparing the peak with the highest intensity (colored with grey illustration in Figure 3.11) to the diluted peaks with lower intensities (shifted to the left and indicated by an arrow in Figure 3.11). Each diluted peak represents a single cell division. Multiple peaks in one sample is explained as multiple populations that are presented in the same sample at a single time-point.

The initial observation was that the cells have been stimulated by the initial ZOL stimulation, as seen by the shift of peaks to the left (from $10^3$) from day 0 to day 3 and day 3 to day 6 (ZOL condition). A second observation is that on day 6 with the ZOL condition, two peaks representing two different populations are present and are denoted as a higher-proliferative population on the left (with greater dilution of cell trace violet dye) and a less proliferative population on the right (with a higher concentration of cell trace violet dye). On day 6, for the ZOL and OKT3 stimulated sample, a positive stimulatory effect of OKT3 is deduced from the higher number of cells with high proliferative characteristics (peak on the left). Moreover, in the ZOL and OKT3 day 6 samples, the low proliferative population (peak on the right) is relatively small compared to the equivalent peak in ZOL-only stimulated sample on day 6. The reduced low proliferative population on day 6 could imply an increase in cell death. It could also mean a decrease in the peak size (peak on the right) of the low proliferative cells caused by the presence of a larger population of proliferating cells (presented with the peaks on the left).

We hypothesize that exhaustion emerges around day 6 following the second stimulation using OKT3 due to the followed decrease in cell numbers observed on day 9. Cell phenotyping using FACS did not show an emergence of a hypoproliferative population on day 6, however, a hypoproliferative population was shown on day 9 and is consistent with exhaustion emergence on Day 6.

*Effect of ZOL/OKT3 stimulation on exhaustion markers (Day 6)*

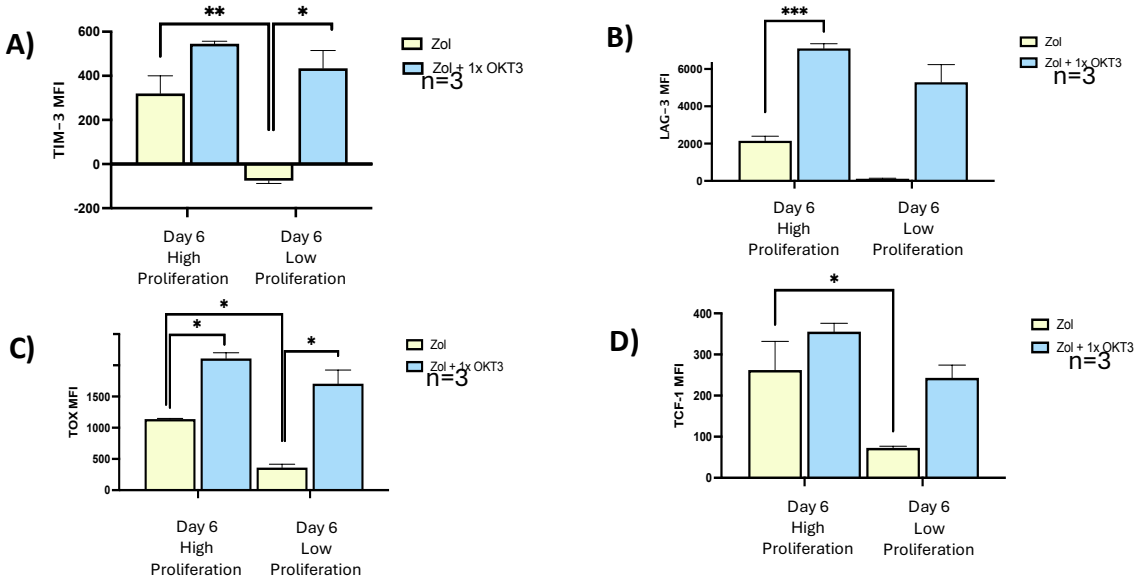



*Figure 3.12: Differential expression of exhaustion markers on high proliferative and low proliferative Vγ9Vδ2 T cells population at day 6 and during a ZOL/OKT3 time-course chronic stimulation model. expression levels (mean fluorescence intensities) of A) TIM-3, B) LAG-3, C) TOX and D) TCF-1 are shown. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

We therefore investigated the two sub-populations of high proliferation (low CTV and peaks on the left (figure 3.11) and low proliferation (high CTV and peaks shown on the right (figure 3.12)) populations (single and double stimulated conditions) emerging on day 6 and compared their levels of exhaustion markers. We looked at levels of TIM-3, LAG-3, TOX and TCF-1 (Figure 3.12). The overall pattern is that there are higher levels of these markers in the population with greater cell trace dilution (those that had undergone more rounds of cell division at day 6), There were similar levels of TIM-3, LAG-3, TOX and TCF-1 on the double stimulated condition (ZOL + 1xOKT3) between low proliferative (high CTV) and high proliferative (low CTV) cells. The double stimulated condition shows higher expression of all markers in the high proliferative population (low CTV) compared to the single-stimulated condition.

The pattern of exhaustion markers expression supports the concept that additional stimulation is associated with the generation of exhausted T cells in cultures. There was a higher expression of exhaustion markers on the subpopulation on day 6, which had gone through greater rounds of cell division, consistent with the hypothesis that cells that received a greater number of stimulations had proliferated more and started to develop a phenotype of exhaustion. The

presumably exhausted population could be responsible for the reduction in cell number observed with multiple stimulations on day 9.

## Effect of Daudi chronic stimulation model on cell numbers and TCR internalization

### *Time-course of Daudi in vitro chronic stimulation model*

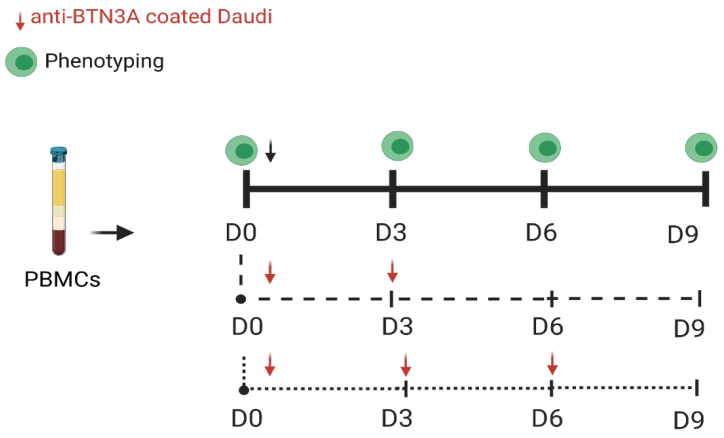



*Figure 3.13: Schematic representation of Vγ9Vδ2 T cells chronic stimulation model in vitro. The stimulus used is Daudi pre-coated with 20.1 anti-BTN3A antibody on Day 0,3 and 6.*

Following the previous two models, we wanted to study how the stimulation response will differ when using a well document target cell to stimulate Vγ9Vδ2 T cells. Daudi cells are a cell line derived from Burkitt's lymphoma and are well-known as efficient stimulators of Vγ9Vδ2 T cells (Cano et al., 2021). In this specific model (figure 3.13), Daudi cells were pre-coated with an anti-butyrophilin 3A antibody (BTN3A known as 20.1 clone) to sensitise them for the recognition of Vγ9Vδ2 T cells by inducing a conformational change of the extracellular domain of BTN3A on Daudi cells (Benyamine et al., 2016). In a different publication, butyrophilin 2A1 was shown to be essential for phosphoantigen reactivity by γδ T cells (Rigau et al., 2020). Co-culture of Daudi with Vγ9Vδ2 T cells was done with an effector-to-target ratio of 10:1 first stimulus, and due to difficulty in re-counting the cells on following time points, the same number of Daudi cells from

the first stimulation was added for a double or triple stimulation. The coating with anti-butyrophilin 3A antibody was done before each time Daudi was added to Vγ9Vδ2 T cells cultures. The stimulations were done on the day of isolation (Day 0), and/or 3 days post isolation (Day 3) and/or 6 days post isolation (Day 6). Each condition (single, double or triple stimulation) of Daudi stimulation is tested for Vγ9Vδ2 T cells exhaustion phenotype and numbers. Figure 3.13 shows that the first stimulus is given at day 0 using Daudi pre-coated with a 20.1 BTN3A clone at a final concentration of 10ug/ml with IL-2.

Then, an additional Daudi pre-coated with 20.1 was added to stimulate cultures on days 3 and 6, with IL-2 added to the cells every 2-3 days. On days 0 (pre-stimulation), 3 (post-first Daudi+20.1 stimulation), 6 (post-first and pre-second Daudi+20.1 stimulation) and 9 (post-all differential stimulations with Daudi+20.1), the phenotypes of differentially stimulated Vγ9Vδ2 T cells were investigated using FACS. Cell numbers of Vγ9Vδ2 T cells are also monitored and counted to study the effects of Daudi stimulation on their expansion.

### *Effect of Daudi chronic stimulation model on cell numbers and TCR internalization*

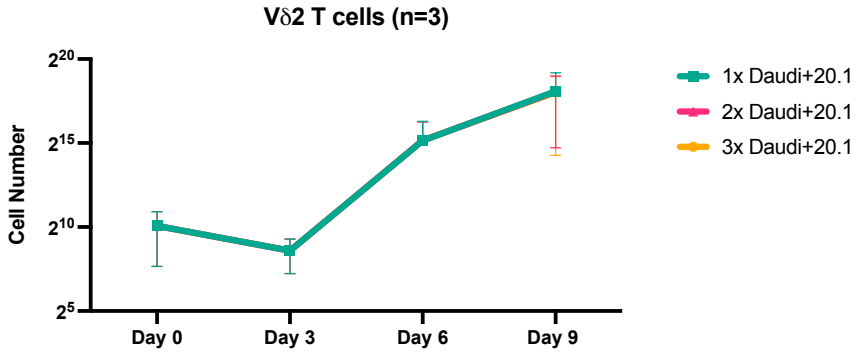



*Figure 3.14: No effects on cell numbers 20.1 precoated Daudi stimulation. PBMCs cultures were stimulated with Daudi cells at 1 to 1 effector-to-target ratio at day 0 and sub proportions were also stimulated again at Day 3 for a second dose or with a third dose at Day 6. Comparisons between cultures only stimulated with single stimulus, double stimulus, and triple stimulus were made at Days 0,3,6 and 9 in regards of Total numbers and are shown for Vγ9Vδ2 T cells. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

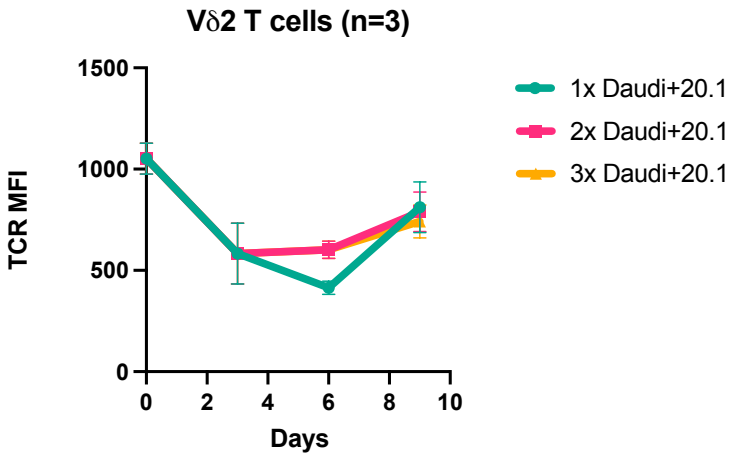



*Figure 3.15: Trends of the T cell receptor (TCR) surface expression levels of Vγ9Vδ2 T cells with increasing stimulation during 20.1 precoated Daudi* stimulation model. *PBMCs cultures were stimulated with Daudi cells at 1 to 1 effector-to-target ratio at day 0 and sub proportions were also stimulated again at Day 3 for a second dose or with a third dose at Day 6. Comparisons between cultures only stimulated with single stimulus, double stimulus, and triple stimulus were made at Days 0,3,6 and. Expression levels were plotted using the mean fluorescence intensities (MFI) of TCR expression on the surface of Vγ9Vδ2 T cells. Statistical analysis was done using a two-way ANOVA and Significance shown for P values <0.05. Three biological replicates were used.*

The absolute cell numbers were counted to measure the impact of chronic activation on T-cell expansion (Figure 3.14). Cell numbers increased proportionally with time in the studied co-cultures, and no stimulation-specific differences were noted on a specific day. In detail, On Day 6, there were no noticeable differences in cell counts between single and double-stimulated conditions as well as on Day 9, where the numbers remained similar regardless of the number of stimulations provided. The findings suggest that three stimulations of Daudi cells pre-coated with 20.1 is insufficient to evaluate functional hyporesponsiveness of Vγ9Vδ2 T cells as suggested by their decreased cell numbers.

The model was also studied further for its effects on the exhaustion of Vγ9Vδ2 T cells, and a selected panel of exhaustion markers were used to check their expression levels on the three conditions tested at different time points. TCR levels were detected to check for any TCR internalization by Vγ9Vδ2 T cells due to Daudi stimulation (Figure 3.15). TCR levels decreased by Day 3 and Day 6 then increased on Day 9 for single Daudi stimulus and decrease by Day 3 then increase on Day 6 and 9 for multiple stimulations which is a similar trend to cell number changes of multiple stimulations conditions, suggesting that there is no TCR internalisation as a

response to Daudi stimulation with this model, as opposed to the OKT3-only model (Figure 3.4B) where TCR levels decrease with increased stimulation in αβ T cells.

*Effect of Daudi chronic stimulation model on the expression of exhaustion markers*

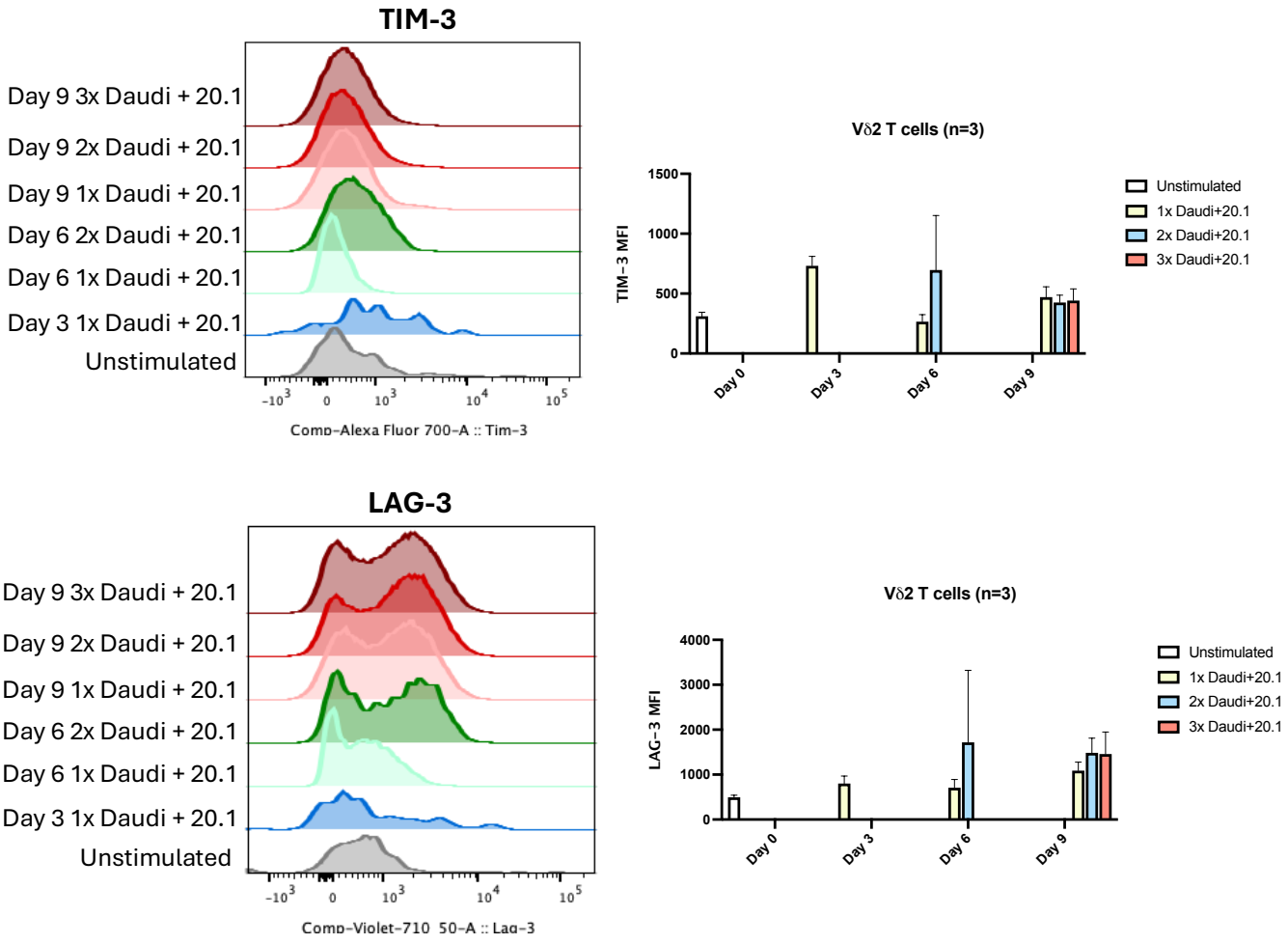



*Figure 3.16: Differential expression of Classical exhaustion markers TIM-3 and LAG-3 on Vγ9Vδ2 T cells during a time-course chronic stimulation model. Expression levels (mean fluorescence intensities) on days 0,3,6 and 9 are shown on the right and histograms of cell counts are shown on the left. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

Exhaustion-associated proteins were also investigated in the study of this Daudi stimulation model. As in the previous model, TIM-3 and LAG-3 were chosen to study the phenotype of Vγ9Vδ2 T cells stimulated with 20.1 coated Daudi cells (Figure 3.16). There was an upregulation of exhaustion/activation markers (TIM-3 and LAG-3) with double-stimulated condition from the baseline (pre-stimulation) on Day 6. The pattern on Day 9 is slightly different, where LAG-3 expression on double and triple-stimulated samples is similar; it is higher in the single-stimulated condition. TIM-3 has no expression differences between all conditions on day 9. For both markers, there is a slight upregulation in the expression on day 3 following the first stimulus. However, the difference in expression is non-significant compared to the ZOL model.

*Effect of Daudi and 20.1 antibody stimulation model on the expression of novel exhaustion markers*

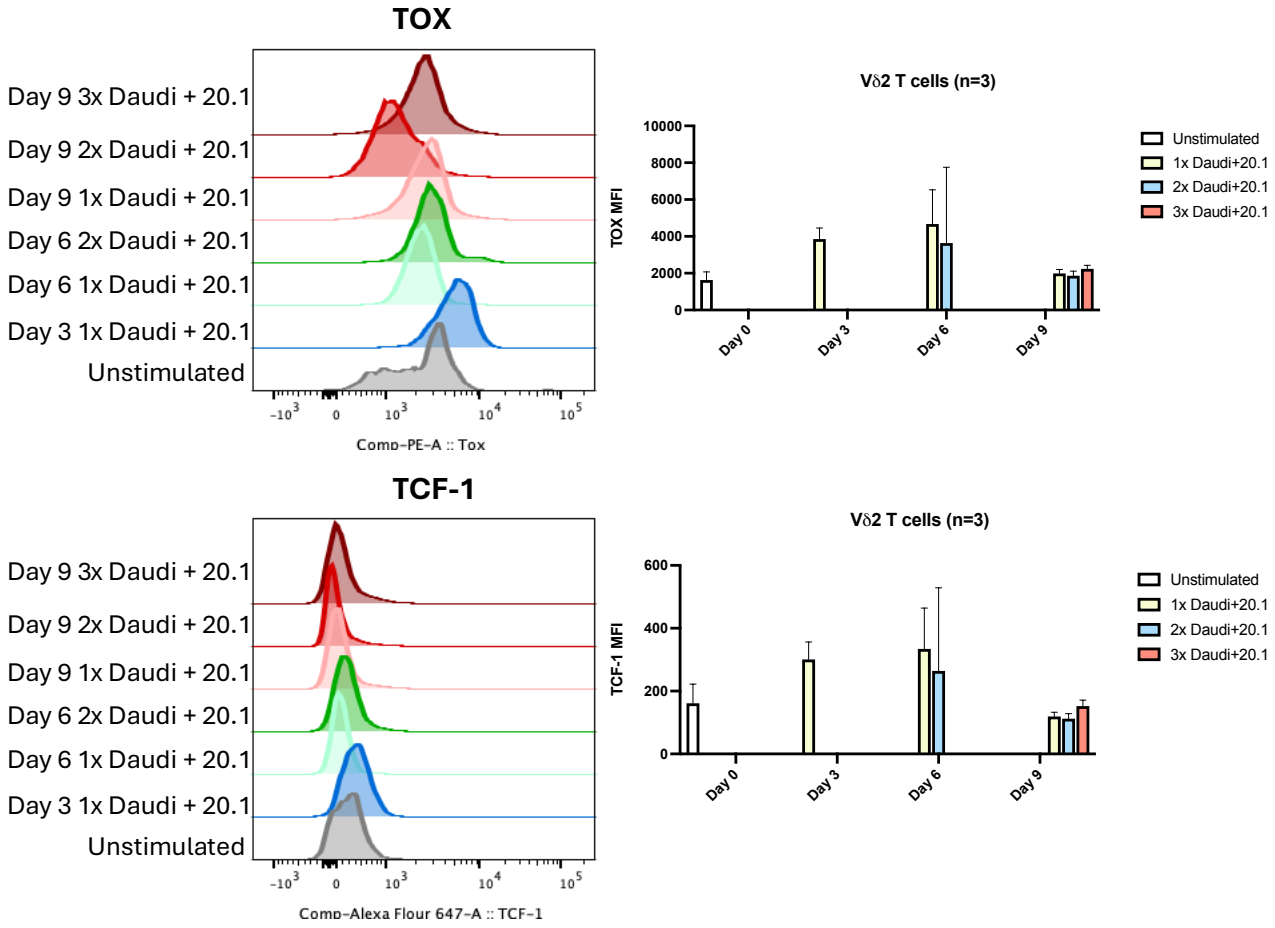



*Figure 3.17: Differential expression of exhaustion markers TOX and TCF-1 on Vγ9Vδ2 T cells during a time-course chronic stimulation model. Expression levels (mean fluorescence intensities) on days 0,3,6 and 9 are shown on the right and histograms of cell counts are shown on the left. Statistical analysis was done using a two-way ANOVA and significance shown for P values <0.05. Three biological replicates were used.*

For novel exhaustion markers in Figure 3.17, an interesting initial observation was that Vγ9Vδ2 T cells showed higher TOX expression levels across time following stimulation, with the highest levels noticed on day 6. The patterns of TOX expression are higher between the baseline (pre-stimulation) condition and day 6. Moreover, both markers are upregulated on day 6 between the single and double stimulated conditions. TOX has been previously reported as an essential marker for exhausted T cells, and it exists at higher levels in terminally exhausted cells. Therefore, the high expression of TOX in the cultures studied under ZOL and OKT3 stimulation

is consistent with the emergence of an exhausted population by day 6. Further validating the existence of an exhausted population, TOX expression patterns also match TIM-3 and LAG-3 expression patterns in the same cultures.

With this model, the patterns of TOX and TCF-1 were almost identical. The expression of TOX and TCF-1 were higher on Day 3 following the first stimulation and on Day 6 for single-stimulated conditions. The expression on double-stimulated samples was not shown to increase on Day 6. In contrast, it decreased marginally. On Day 9, the levels of both expression markers (TOX and TCF-1) were less than previous time points and were almost equal for single, double and triple-stimulated samples.

The Daudi model does not associate with differences in exhaustion/activation markers expression levels in the different conditions tested. TIM-3, LAG-3, TOX and TCF-1 expression levels did not show variations between differentially stimulated conditions, suggesting no significant differences in the cell's phenotype is detected in contrast with the ZOL model where significant differences in exhaustion marker expression were highlighted in the double stimulated Vδ2 T cells compared to single-stimulated conditions. Moreover, the cell numbers did not represent reduced or significant changes with multiple stimulations using Daudi cells, where in the ZOL model, the Vδ2 T cell numbers varied significantly and were reduced with multiple stimulations.

### *Effect of Daudi chronic stimulation model on cell Proliferation*

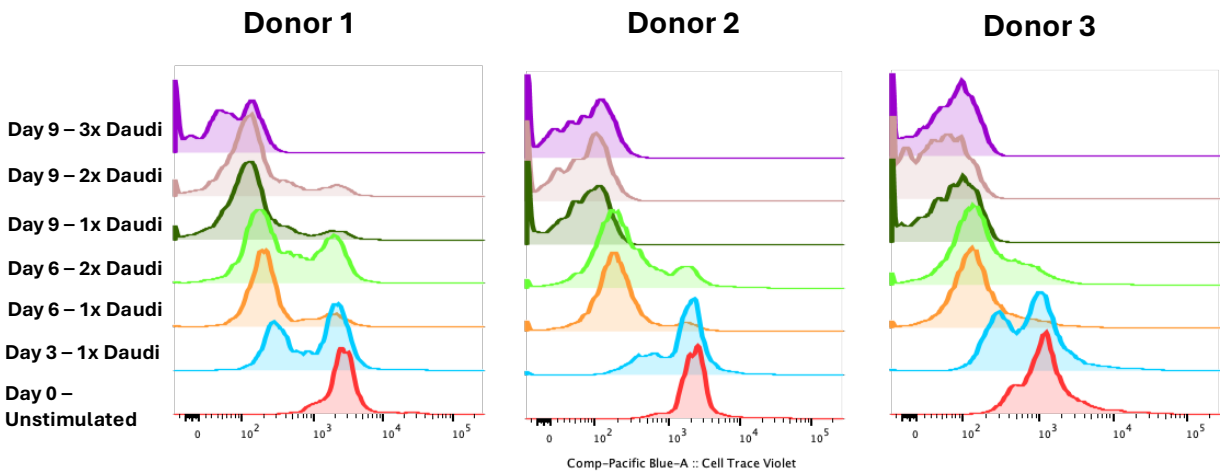



*Figure 3.18: Proliferation capacity of Vγ9Vδ2 T cells during 20.1 precoated Daudi stimulation model. The figures show proliferation of cells using a proliferation dye (cell trace violet). the assay is analyzed by comparing the peak with the highest intensity (colored with grey illustration) to the diluted peaks with lower intensities (shifted to the left and indicated by an arrow). Each diluted peak represents a single cell division. Multiple peaks in one sample is explained as multiple populations that are presented in the same sample at a single time-point. Three biological replicates were used.*

Again, the proliferation of cells during the time-course study and with different stimulations was studied using a cell trace dilution assay (Figure 3.18). First, the stimulation of the cells by 20.1 monoclonal antibody pre-coated Daudi cells is validated by the appearance of a second peak on Day 3 (between $10^3$ and $10^2$). The second observation is that on day 6, with the single stimulus condition, the peak on the left increased and correlated to more activation. Interestingly, for the double stimulus sample on day 6, the peak on the right has increased and represents a low proliferative population (peak at $10^3$) with a high concentration of cell trace violet dye, similar to what was observed with the ZOL model on Day 6 proliferation of Vγ9Vδ2 T cells following double stimulation (ZOL and OKT3). On Day 9, there is an evident disappearance of the low proliferative population for two donors and an apparent increase in the high proliferative T cells population (Peak at 0). With longer duration, the low proliferative populations from day 6 failed to persist, as shown by the high proliferation of cells on day 9 with all conditions.

As described before with the ZOL model, the disappearance of a low proliferative population on day 9 could imply either increased cell death or the population is overshadowed by a higher number of proliferating cells. In this model, there is an increase in the low proliferative population on day 6 after the second stimulation, however, as shown with the exhaustion markers expression profiles, it does not correspond to the exhaustion phenotype as demonstrated by the ZOL model. The results might indicate the presence of heterogeneous populations but not necessarily an exhausted population. To use this model for exhaustion and hypofunctionality evaluation in Vγ9Vδ2 T cells, further optimisation of the co-culture effector-to-target ratios and duration of stimulation in culture are required.

## Conclusion

Three different stimulation models of chronic stimulation were evaluated for indication of phenotypic and functional exhaustion in Vγ9Vδ2 T cells. Exhaustion is a state of functional hyporesponiveness in T cells. Cell counts during stimulations can be used as an indicator of Vγ9Vδ2 T cells hyporesponsiveness along the duration of the model. Expression of classical exhaustion markers such as TIM-3 and LAG-3 are also used to evaluate exhaustion in T cells. Yet, these

markers are also considered activation markers in T cells and do not exclusively define exhaustion. Therefore, additional markers TOX and TCF-1 are also detected as they are well known for their association with initiation of exhaustion in conventional T cells and high expression in exhaustion precursor cells, respectively. TCF-1 is a general marker of progenitors and also reported to be preferentially expressed in early memory stem like T cells. With chronic stimulation, exhaustion emerges in T cells, resulting in increased TOX expression and low levels of TCF-1.

**T cell numbers**

The ZOL activation model led to higher differential expansion with single and multiple stimulations of Vγ9Vδ2 T cells compared to the other models evaluated. A striking difference in reduced Vγ9Vδ2 T cell numbers between single and multiple stimulations provided strong evidence of hyporesponsiveness following the second and third stimulation. However, the reduction in numbers from the Zol model could be affected by Activation- induced cell death was not validated during the study and would potentially aid in understanding if the cell ' s response is solely from chronic stimulation. Moreover, elongating the timeline of the study would have disclosed the later effects following the fall in numbers on day 9. The TCR expression response was surprising as Vγ9Vδ2 T cells did not experience reduced TCR levels following Zol and OKT3 stimulations, which could indicate that they respond differently to conventional T cells or that they are not chronically stimulated from the timeline used and a longer stimulation or the use of different OKT3 concentrations might lead to different responses that were not considered in this experiment. The OKT3 model was associated with a trend of decreasing cell numbers in Vγ9Vδ2 T cells but this effect size did not reach statistical significance. The OKT3 model was associated with TCR expression decrease with increased stimulations in conventional T cells and could have affected the reliability of cell counts at the times shown. Considerations of methods that can prevent TCR internalisation prior to analysis would have increased the confidence in the study and the comparisons in terms of counts of conventional T cells versus Vγ9Vδ2 T cell counts. Moreover, as OKT3 was not commonly used for Vγ9Vδ2 T cell activation, it would have been informative to optimise the concentrations of OKT3 used and the length of stimulation model to validate the optimal conditions for chronic stimulation. The Daudi model did not associate with cell numbers decrease trend with increased stimulations. Daudi cells were selecte as a stimulant based on their previously reported use as targets for Vγ9Vδ2 T cells stimulation. The wells used for co-culture

were coated with 20.1 antibody to enforce the stimulation of Vδ2 cells TCR via the BTN3A1 and BTN2A1 receptors. However, the effector-to- target ratios studied might not have beenoptimal to chronically stimulate Vγ9Vδ2 cells. When comparing the Daudi model to the Zol model, it seems like the response to Daudi did not challenge the cell expansion over the same period used with the Zol and OKT3 stimulations, stressing the fact that further variations to the amount of Daudi used or even comparison to a model where only Daudi (without 20.1 pre-coating) could have confirmed the efficiency of the model developed to study Vγ9Vδ2 T cells hyporesponsiveness or exhaustion.

**Proliferation (CTV dilution assay)**

In the ZOL model, cell trace violet experiments showed two populations in terms of cell trace violet dilution (indicating previous proliferation since day 0) on day 6 in single-stimulated and double-stimulated samples. The populations were termed high proliferative and low proliferative and were phenotyped for their expression of exhaustion markers (TIM-3, LAG-3, TOX and TCF-1). The expression of exhaustion markers on both high proliferative and low proliferative populations was higher on the multiple stimulated conditions than the single stimulated. Exhaustion markers were also higher in the cell trace dim sub-populations that have gone through more replication cycles compared to the cell trace high sub-populations. The observed behaviour of the cell trace high subpopulation is consistent with a model where cells receiving more stimulations result in lower proliferation and upregulation in exhaustion markers—additional evidence of the ZOL model to be used for exhaustion phenotyping. We hypothesised that the cell trace high-exhaustion markers high subpopulation on day 6 following the second OKT3 stimulation represents cells that have stopped proliferating, hence, the decreased total cell numbers observed in the multiple stimulated conditions on day 9. There is a possibility that dead cells by day 9 are not shown for proliferation and might be of an exhausted phenotype or cells that experienced activation-induced cell death, therefore disappeared from the culture and are not detected with proliferation assays.

The Daudi model did not show significant differences in exhaustion markers expression between single and multiple stimulated conditions globally on a whole population level. Hence, it will not be chosen as a suitable model for exhaustion evaluation in Vγ9Vδ2 T cells. There was an indication of the formation of a heterogenous double population from the cell trace violet dilution study suggesting a subpopulation with reduced proliferation. However, there were no significant

differences in cell number between conditions with different numbers of Daudi challenges as was shown with the Zol Model. Further optimisation to the model or the use of different markers of proliferation might inform more differences between the conditions tested. Moreover, the inclusion of an unstimulated control tracking for all time points would have shown differences in proliferation and insights into the effects of chronic stimulation on cell proliferation with all models evaluated. The OKT3 model did not include any proliferative studies but could have greatly benefited from proliferation tracking and differences shown in Vγ9Vδ2 T compared to conventional T cells.

**Global TIM-3 and LAG-3 expression**

Expression of TIM-3 and LAG-3 were dramatically increased in multiple stimulated Vγ9Vδ2 T cells compared to single stimulated condition using the ZOL model. This was not recapitulated with the OKT3 or Daudi models of stimulations. The increase in TIM-3 and LAG-3 is not unexpected following stimulation of T cells since they are activation as well as exhaustion markers. Their lack of expression in the OKT3 and Daudi models further indicates the failure of the models to stimulate the T cells with the shown timeline and at the chosen concentrations of OKT3 and amount of Daudi cells used. Since TIM-3 and Lag-3 are also activation markers, their expression does not credibly prove exhaustion emergence in the studied populations. The inclusion of another marker, such as PD-1, would have added more confidence in the levels or markers tested, but it was not possible for the technical failure of the antibodies to be detected by the FACS.

**Global TOX expression**

TOX was used to validate the differences observed with TIM-3 and LAG-3. Using the ZOL model, TOX expression was significantly increased in the double-stimulated condition on day 6. TOX expression was also reported to be expressed in exhausted conventional T cells on Day 6, which is consistent with the notion that the Zol protocol could successfully chronically activate Vδ2 cells and that the populations emerging on day 6 having an exhausted phenotype. Hence, the ZOL model can be further used to identify hypofunctional and phenotypic exhaustion in Vγ9Vδ2 T cells.
The Daudi stimulation model did not affect single or multiple stimulated Vγ9Vδ2 T cells. The OKT3 model was not tested for TOX expression, and it would have been more informative if it had been included in the analysis.

**Global TCF-1 expression**

TCF-1 levels had a similar trend to TOX; however, the increased level is much lower than TOX, as indicated by the MFI values shown in Figure 3.11. TCF-1 expression also indicates the presence of TCF-1 high progenitor cells within the heterogenous population in addition to the population with high TOX levels on day 6. Even though exhaustion definitions are not fully understood in conventional T cells, Increased TCF-1 levels were also reported in αβT cells and linked with progenitor-exhausted T cells that are not terminally exhausted and are highly desirable for sustained effector functionality. This significant change in TCF-1 between single and double-stimulated conditions is very interesting and an indicator of the presence of progenitor cells in addition to cells that express high levels of TOX, which is more generically associated with exhausted cells. The results also show an exhaustion program initiation in Vγ9Vδ2 T cells and display relevant TOX-TCF-1 dynamics in Vδ2 T cells, amplifying the superiority of the ZOL model to the other models tested.

However, as the ZOL model might not have resulted in a fully exhausted population but a partially exhausted population, it makes the analysis of the exhausted cells only in terms of functionality, proliferation, or marker expression generally more challenging.

# Chapter 4

## Functional studies using single and double-stimulated Vδ2 T show overlapping responses to various cancer targets.

### Introduction

Previous experiments identified phenotypic and transcriptomic differences between the differentially stimulated Vδ2 T cell populations. Functional experiments are required to gain a more profound understanding of the effects of gene expression differences on Vδ2 T cell functionality. In this chapter, functional experiments were designed to assess the differences in functionality and exhaustion markers expression of single and double-stimulated cells in response to stimulation with different target cells. The first experiment evaluated the proliferation and survival of Vδ2 T cells when co-cultured with Daudi target cells at different effector-to-target ratios. The second experiment was designed to study the emergence of hypo-functionality and heterogeneity of Vδ2 T cell populations following the dual stimulation with OKT3. The final experiment assessed the response of Vδ2 T cells following double stimulations by cell counts and expression of exhaustion markers (PD-1 and LAG-3) in response to re-challenge with different targets with different levels of stimulating potential to Vδ2 T cells (i.e. Daudi, THP-1 and Jurkat) on alternating days.

## *Co-culture using Daudi and Vδ2 T cells*

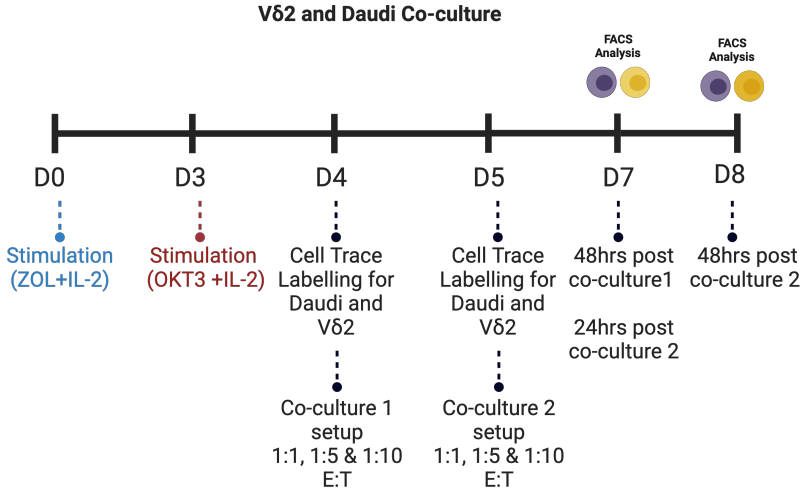



*Figure 4.1: A representation of Daudi and Vδ2 co-culture following single and double stimulation of Vδ2 T cells.*

## *Co-culture using Daudi and Vδ2 T cells following 1 Day of OKT3 stimulation*

A co-culture of Vδ2 T cells against target cells was performed to investigate the functionality of single (ZOL) and double-stimulated (ZOL+OKT3) cells. The cells were expanded using ZOL on day 0 and followed by OKT3 on day 3, with IL-2 provided every 2-3 days. Following OKT3 addition on day 3, cells were labelled with cell trace dyes to track their proliferation in the co-culture. Vδ2 T cells were labelled with cell trace violet (CTV), and target Daudi cells were labelled with cell trace yellow (CTY). The single and double-stimulated cells were stained and counted 24 and 48 hours after co-culture. The co-culture was also tested with the different effector-to-target ratios of 1:1, 1:5 and 1:10 Vδ2 : Daudi ratios (Figure 4.1).

The cells were stained at Day 0, and representative histograms are shown in Figure 4.2. These demonstrate efficient cell trace labelling of Vδ2 T cells (with CTV) and Daudi (with CTY), and indicate no significant proliferation at the time of co-culture start. The cells (single or double stimulated) at 1:5 and 1:10 behaved similarly, and all Vδ2 T cells proliferated after 48 hours of target addition shown by the peaks shift to the left when compared to the control (grey peaks) (Figure 4.3). The cells with a 1:1 ratio show two distinct peaks indicative of 2 different

populations with different proliferative patterns. Daudi cells also show signs of proliferation with the 1:1 ratio only. There are no differences in cell counts between single and double-stimulated Vδ2 T cells as they show similar counts even with different ratios of target cells (Figure 4.4). However, the Vδ2 T cells only show a reduced cell count on the double-stimulated condition, consistent with what was observed previously in Chapter 1. Of interest, in Daudi's presence, with the 1:1 co-culture there a rescue of the double-stimulated Vδ2 T cell proliferation shown by increased numbers compared to the Vδ2 only control. There is no difference in T cell numbers between double and single stimulations with 1:5 and 1:10 ratios.

It was also noticed that Daudi cell numbers increase in the co-culture in response to Vδ2 compared to Daudi only control. From the cell counts (Figure 4.4), there seem to be more Daudi cells in culture than Vδ2 T cells for all ratios tested (1:1, 1:5 and 1:10) on Day 4 (48hrs following the addition of Daudi cells). Even though Daudi cells show an apparent effect on reducing Vδ2 cell numbers, it doesn't seem to affect Vδ2 cells proliferation. The results show that the selected time points were insufficient to look at noticeable differences in the functionality or proliferation of single or double-stimulated Vδ2 T cells.

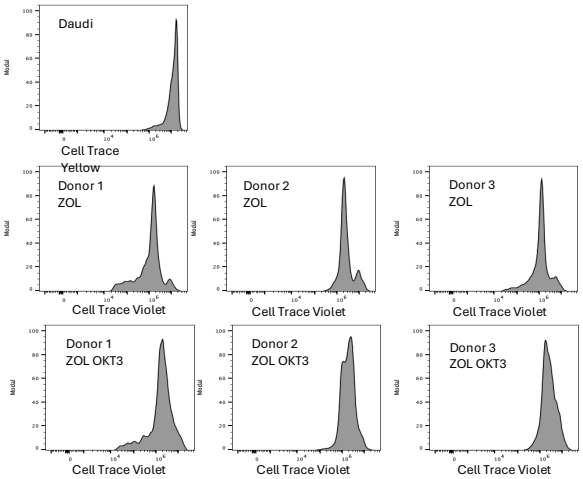



*Figure 4.2: Cell Trace labeling of target cell (Daudi) and Vδ2 T cells (single stimulated and double stimulated) at day 0.*

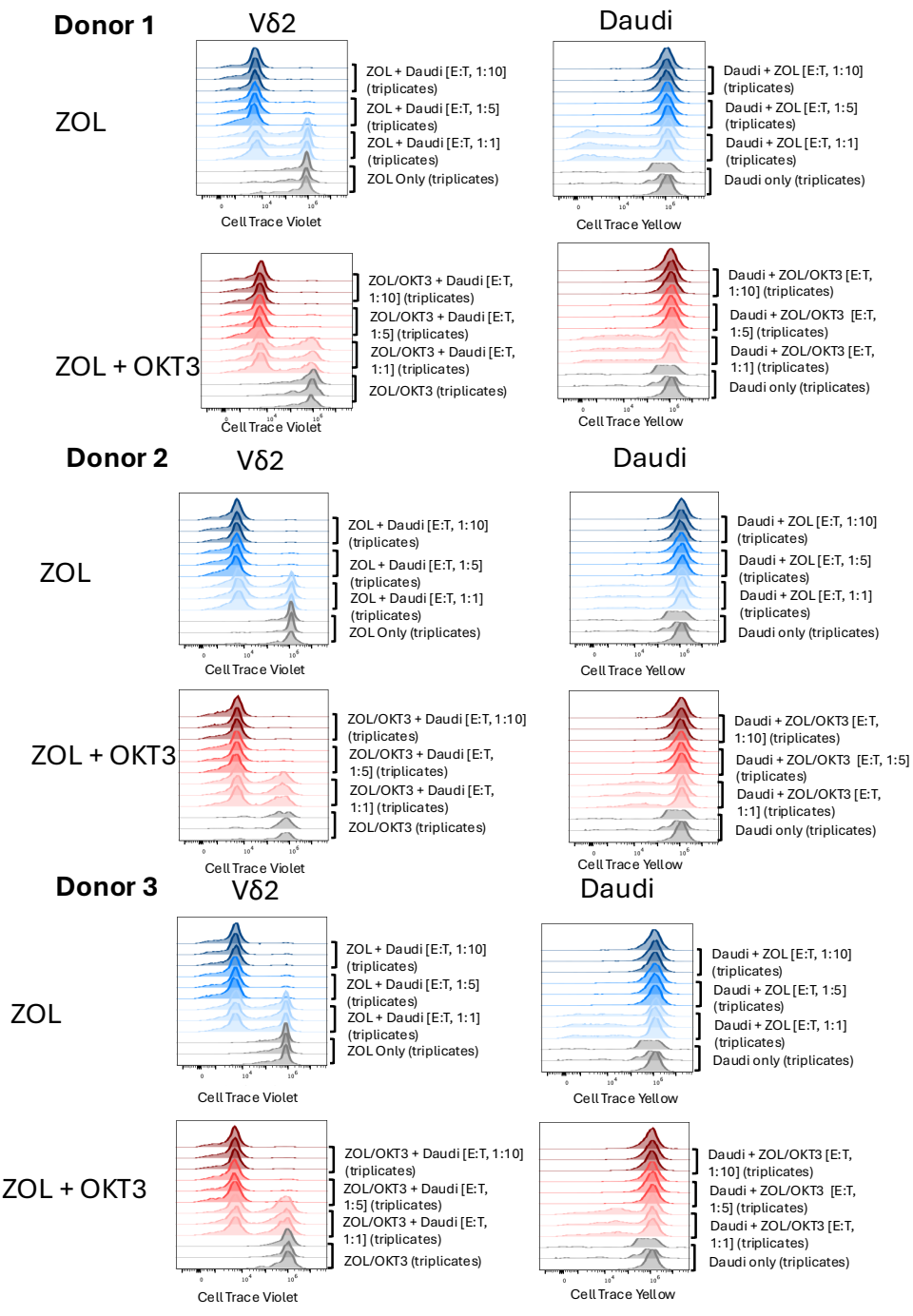



*Figure 4.3: Histograms of Cell Labelling of Vδ2 T cells and different ratios of Daudi (1:1, 1:5 and 1:10 Vδ2:Daudi) after 48 hrs co-culture with single stimulated or double stimulated Vδ2 T cells conditions. The co-culture was set on day 4 (1 day after second stimulus was added. Triplicates of Three biological donors were used.*

## Cell Counts

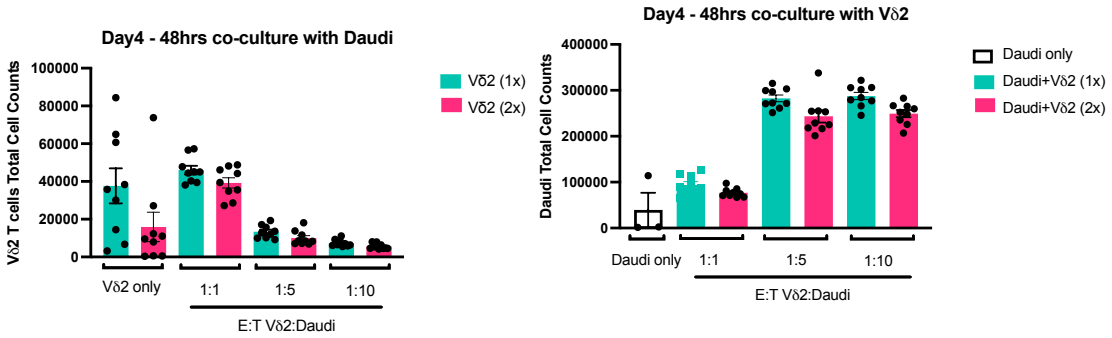



*Figure 4.4: Cell counts of Vδ2 T cells (single and double stimulated) and Daudi cells following 48 hrs co-culture. The co-culture was set on day 4 (1 days after second stimulus was added to Vδ2 T cells). All statistical comparisons between Daudi with Vδ2 (1x) and Daudi with Vδ2 (2x) were non-significant. Triplicates for three biological donors were used.*

### *Co-culture using Daudi and Vδ2 T cells following 2 Days of OKT3 stimulation*

A further co-culture experiment was next performed. The changes were as follows: the target cells (Daudi) were plated with Vδ2 T cells after two days of OKT3 addition as opposed to 1 day after OKT3 stimulation in the previous experiment (Figure 4.1). This change was to detect if there are enhanced effects on Vδ2 T cells following 48 hours of OKT3 stimulation. The effector-to-target ratios were 1:1, 1:5 and 1:10 as before; the total number of Targets and Vδ2 T cells plated remained unchanged from the previous experiment. In summary, for the 1:1 ratio, 25000 Vδ2 T cells were plated with 25000 Daudi cells and adjusted to 10000 Vδ2 T cells and 50000 Daudi cells for 1:5 and 5000 Vδ2 T cells with 50000 Daudi for the 1:10 effector-to-target ratio. Figure 4.5 shows the cell labelling with cell trace for Vδ2 T cells and Daudi at day 0 before setting the co-culture. This confirms that the cells are correctly labelled, and any proliferation differences will be detected accurately. It is shown that there is more proliferation of Daudi cell cultures with Vδ2 T cells at the 1:1 effector-to-target ratio compared to the other ratios. Comparing Vδ2 T cells with single and double stimulations after 24 or 48 hours (Figure 4.6 and Figure 4.7), the histograms of the double-stimulated cells indicate that there are more Vδ2 cells in this population than the single stimulated. However, both differentially stimulated Vδ2 T cells still show two different proliferative populations with the 1:1 co-cultures compared to the 1:5

and 1:10, where complete proliferation is shown by the cells and represented by peak shifts to the far left.

Both single and double stimulated samples show incomplete proliferation of Vδ2 cells with the 1:1 ratio, as indicated by the double peaks in Figures 4.6 and 4.7. This could be either a result of using a lower number of Daudi cells in the 1:1 (25000:25000 cells E: T) ratio compared to 1:5 (10000:50000 cells E: T) and 1:10 (5000:50000 cells E: T) effector-to-target ratios, which leads to lower stimulation of Vδ2 T cells, and this correlates with the fully proliferated Daudi cells in the same 1:1 culture. Alternately, the differential effect could be related to the use of higher Vδ2 T cell numbers in the 1:1 ratio compared to 1:5 and 1:10 as the numbers varied according to the ratios required.

Single-stimulated Donor 1 is associated with reduced Daudi cells at the 1:1 ratio compared to double-stimulated. However, the other two donors show a different pattern. The results indicate the presence of high variability between donors and the need to test more time points to identify the best window to measure the difference in cell functionality in Vδ2 T cells against target cells. The cell counts (Figure 4.8) show an inversely proportional relationship between Vδ2 T cells and Daudi at 24 and 48 hours after co-culture, as with increased Ratios of Daudi cells, there is a decrease in cell numbers of Vδ2 T cells (single or double stimulated). It would be interesting to investigate this in the future by looking at the functionality of Vδ2 T cells, such as killing in response to Daudi stimulation and understanding the effects of Daudi on the Vδ2 cell numbers shown.

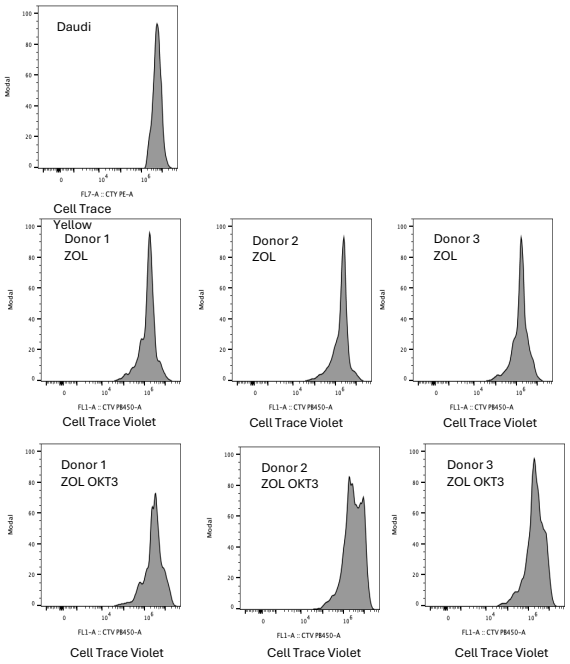
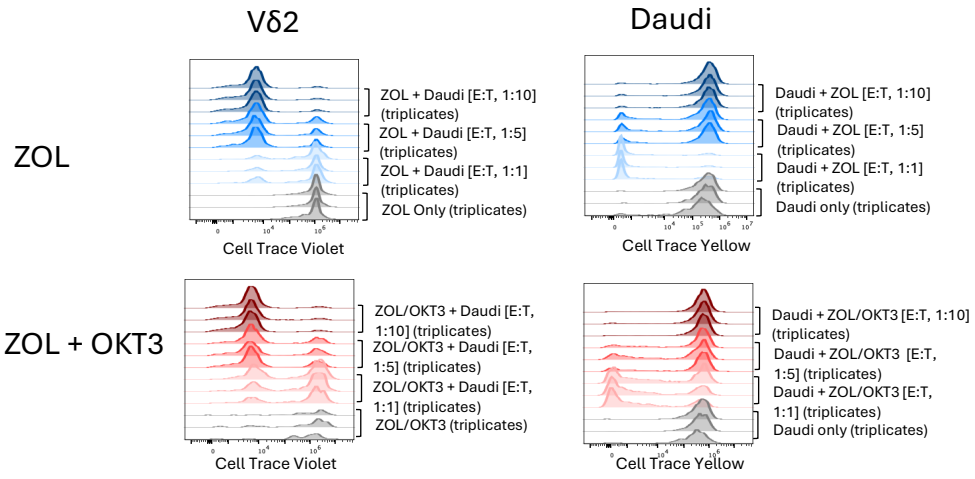
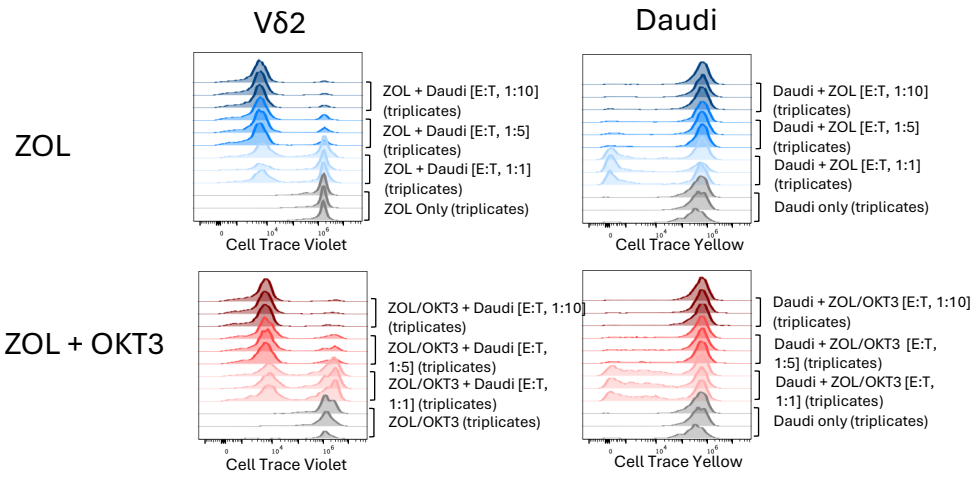



*Figure 4.5: Cell Trace labelling of the target cell (Daudi) and Vδ2 T cells (single stimulated and double stimulated) at day 0.*

**Donor 1**



**Donor 2**




**Donor 3**

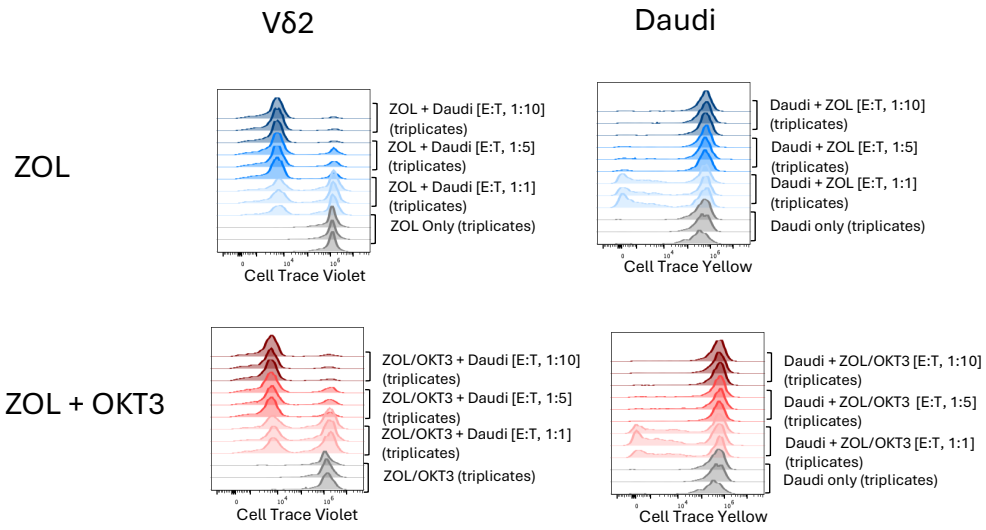
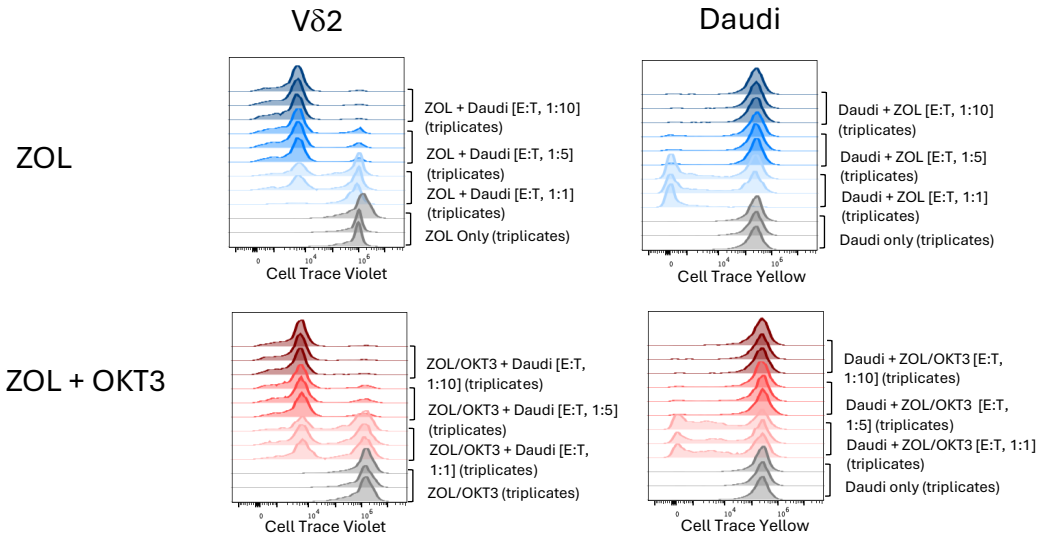
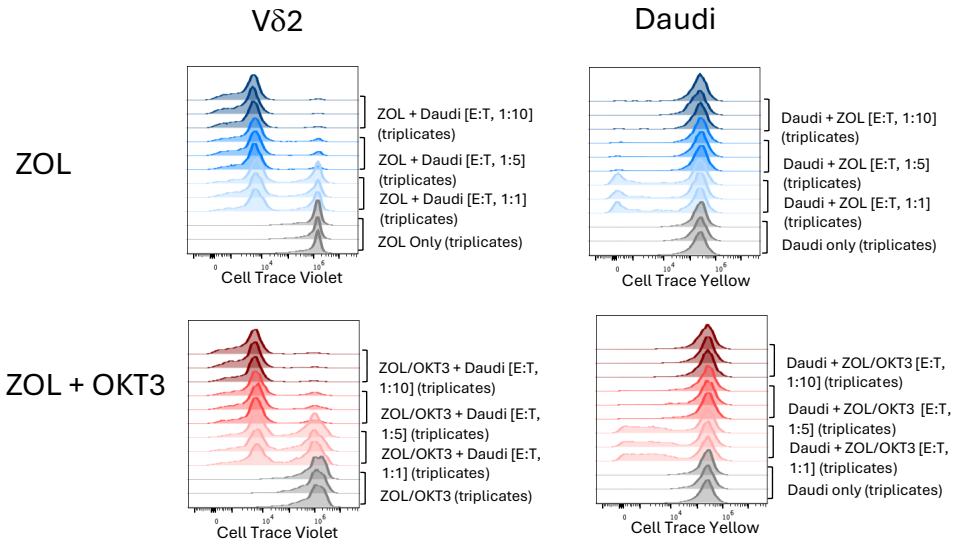



*Figure 4.6: Histograms of Cell Labelling of Vδ2 T cells and different ratios of Daudi (target cells) after 24 hrs co-culture with single stimulated or double stimulated Vδ2 T cells conditions. The co-culture was set on day 5 (2 days after second stimulus was added to Vδ2 T cells). Triplicates for three biological donors were used.*

**Donor 1**




**Donor 2**



**Donor 3**

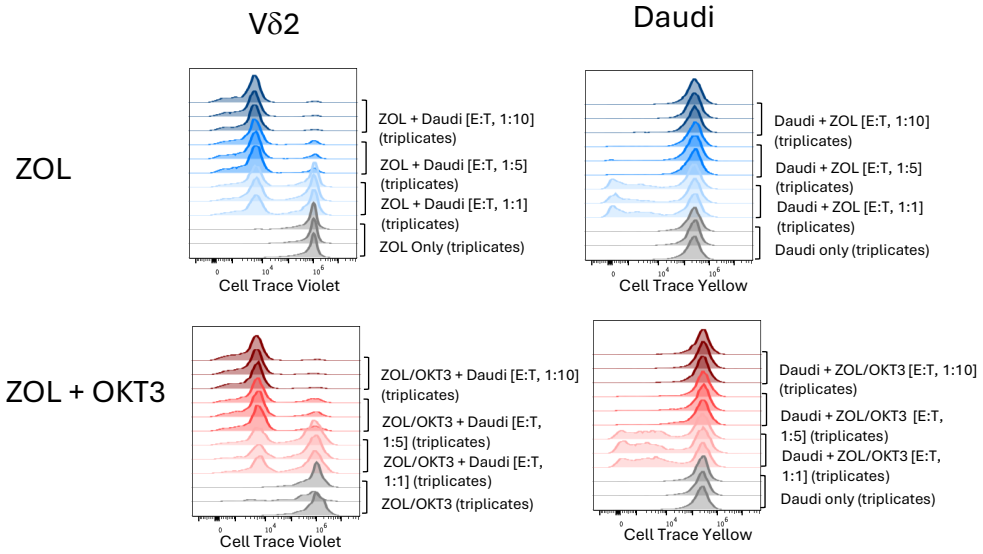



*Figure 4.7: Histograms of Cell Labelling of Vδ2 T cells and different ratios of Daudi (target cells) after 48 hrs co-culture with single stimulated or double stimulated Vδ2 T cells conditions. The co-culture was set on day 5 (2 days after second stimulus was added to Vδ2 T cells. Triplicates for three biological donors were used.*

## Cell Counts

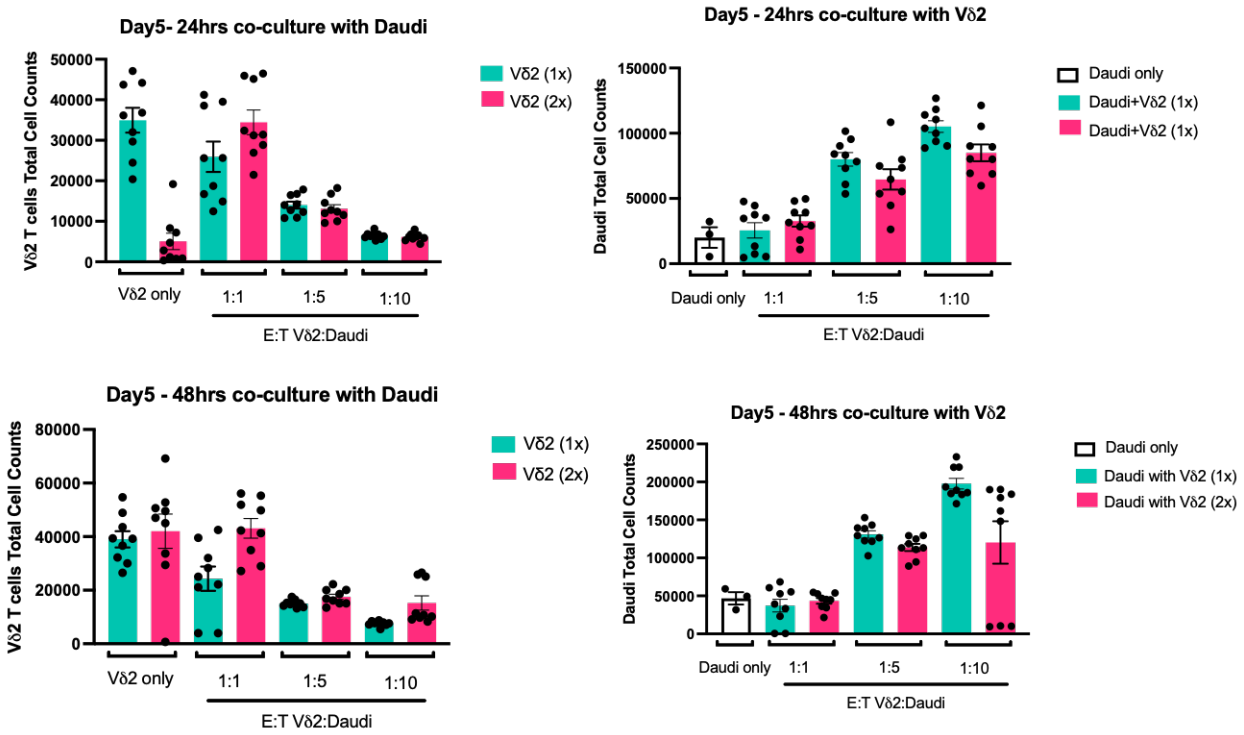



*Figure 4.8: Cell counts of Vδ2 T cells (single and double stimulated) and Daudi cells following 24 and 48 hrs co-culture. The co-culture was set on day 5 (2 days after second stimulus was added to Vδ2 T cells). All statistical comparisons between Daudi with Vδ2 (1x) and Daudi with Vδ2 (2x) were non-significant. Triplicates for three biological donors were used.*

### CFSE tracking in Vδ2 T cells

Previously, we hypothesised that a hypo-proliferative population of Vδ2 T cells emerges on day 6 following double TCR stimulation to account for the reduction in T cell numbers following the second stimulation. We then wanted to study when hypo-proliferative cells emerge following dual stimulation with ZOL and OKT3 stimulation. An experiment was designed with cell trace labelling of Vδ2 T cells at different time -points after OKT3 addition. The cells are then tracked daily from the labelling day and checked for proliferation differences between the single and double-stimulated samples. A schematic of the cell labelling of 24 hours of OKT3 stimulation and 48 hours of OKT3 stimulation are shown in figures 4.9 and 4.10, respectively.

The aim of this experiment was to identify which timepoint following the second stimulation (OKT3) demonstrated a reduction in the cell trace violet dilution compared with the ZOL alone conditions. In the figures, the double stimulation is depicted in red, whilst histograms from single stimulation conditions are depicted in blue. Hence, a red histogram to the right of a blue histogram at any timepoint would be indicative of relative hypoproliferation as a result of the second stimulus.

In the first experimental approach, CTV labelling was performed 24 hours following OKT3 stimulation (Day 4), and the cells were followed for three days after as depicted in Figure 4.9. The figure shows three donors with proliferation plots of single (ZOL) and double (ZOL + OKT3) stimulated samples. On day 4, the labelling starting point is shown and indicated by a peak at the far right. With proliferation, the cells' peaks move to the left, which displays proliferation. In the alternate approaches, CTV labelling was performed on Day 5 (Figure 4.10), and the data shows that single stimulated cells proliferate in donors 1 and 2. The double-stimulated cells are also proliferating and seem higher in donors 1 and 3. A heterogeneity in the double-stimulated populations has been observed and is shown in 2 out of 3 donors with double peaks at the analysis time. Moreover, it is essential to note that responses from different donors were inconsistent and showed differences in responses due to donor variability, commonly known as a limitation of studies on Vδ2 T cells (Burnham et al., 2020).

Day 4 to Day 7 (1 Day after OKT3 stimulation)

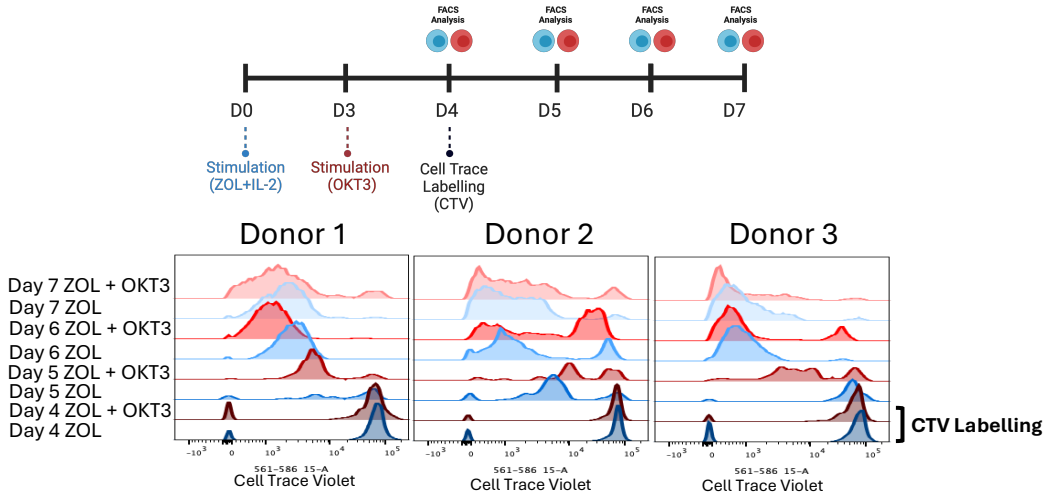



*Figure 4.9: Proliferation tracking of Vδ2 T cells stained on Day 4 and followed until day 7. Three biological donors are shown and two condition of stimulations (ZOL and ZOL +OKT3).*

Similar observations were shown when the cells were stained on day 5 (two days after OKT3 stimulation), with low proliferation differences (Figure 4.10). Still, on day 7, the double-stimulated sample seemed to have a higher low proliferative population for one donor (donor 2). Moreover, donor 2 showed lower proliferation on day 6 with the double-stimulated sample as the shift is slightly lower than the single-stimulated (ZOL) cells.

Day 5 to Day 7 (2 Days after OKT3 stimulation)

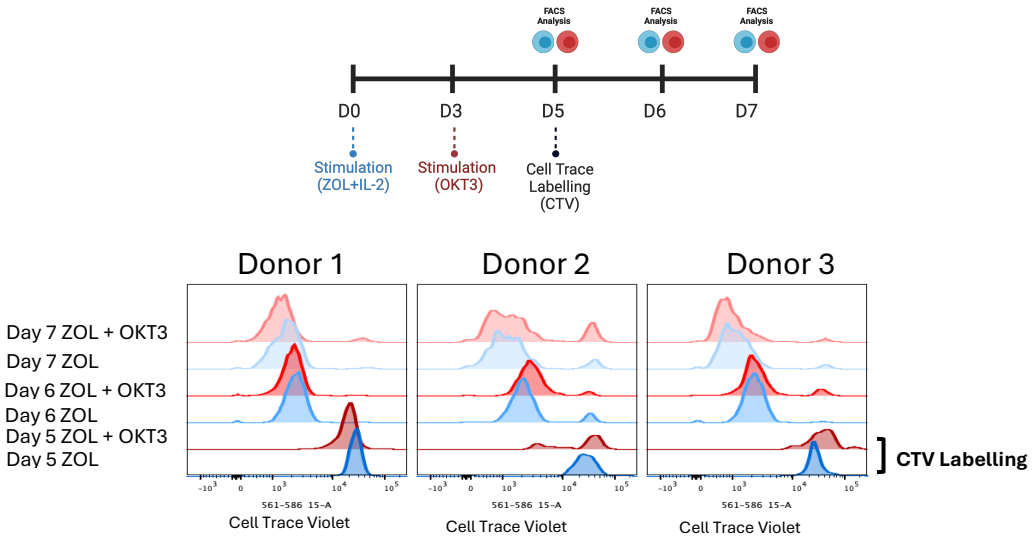



*Figure 4.10: Proliferation tracking of Vδ2 T cells stained on Day 5 and followed until day 7. Three biological donors are shown and two condition of stimulations (ZOL and ZOL +OKT3).*

In conclusion, a consistent pattern of CTV peak shift to the left was expected with the double-stimulated samples of all donors if exhaustion hypofunction was the cause of reduced cell numbers with chronic stimulation. However, this was not observed at any time, suggesting that the reduction in cell numbers following OKT3 is not due to reduced cell proliferation. In contrast, the drop in cell numbers observed with the double-stimulated cells may be due to the poor survival of cells after the second stimulation. Another reason could be the population heterogeneity where proliferation could represent multiple developmental states such as terminal effector T cells and cells experiencing activation-induced cell death and, therefore, disappearing

at specific time points. Preferential proliferation to OKT3 instead of ZOL could be another possibility of what is shown in this study, whereby cells that were not adequately stimulated from ZOL, proliferate more following OKT3 stimulation. Exhausted T cells could also be represented within the low proliferative population (with high CTV). Cell Trace dilution experiments did not identify significant differences for short-term proliferation in Vδ2 T cells. However, it is critical to consider that the results of this study were snapshots of the cells at specific times and do not consider the history and dynamics of the proliferation of Vδ2 T cells between the selected time points.

## Re-stimulation assay of Vδ2 T cells with target cells

A potential readout of hypofunctionality and a sign of T cell exhaustion is the inability of T cells to engage with target cancer cells and the failure to initiate cytotoxicity, cytokine release or proliferative responses. To evaluate this exhaustion characteristic, target cells were added to single or double-stimulated Vδ2 T cells and read outs of exhaustion/activation markers expression levels and cell counts were performed.

The experiment was designed to study if hypofunctionality emerges with the double simulated Vδ2 T cells following a re-challenge with chosen targets added on alternate days (shown in Figure 4.11).

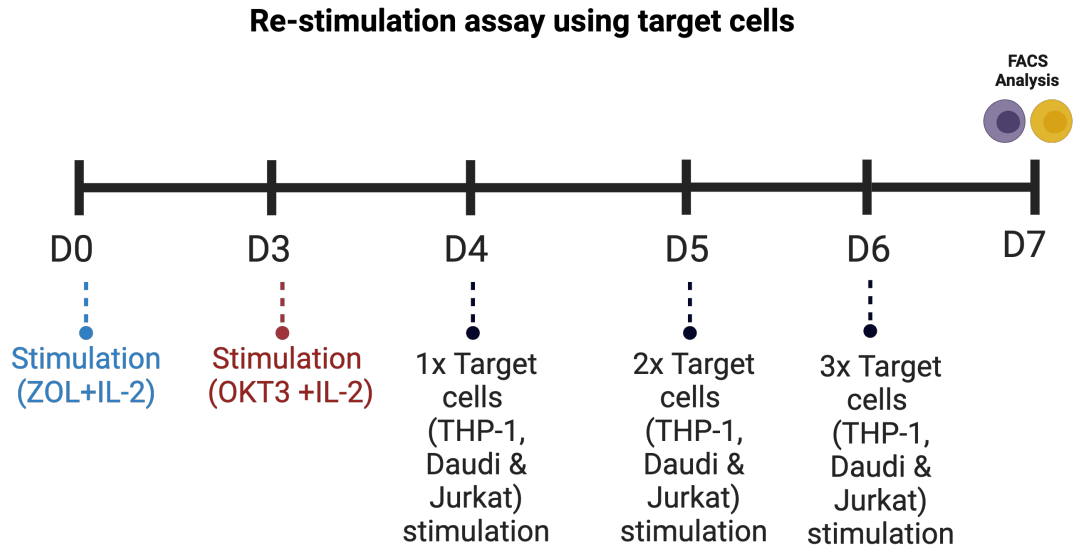



*Figure 4.11:: A representation of the restimulation assay using target cells on single and double stimulated Vδ2 T cells*

Different target cells, including Daudi, THP-1 and Jurkat cell lines, were chosen to provide a range of sensitivity to killing by Vδ2 T cells. Daudi cells are a B cell lymphoma and are highly sensitive and a well-known target for Vδ2 T cell stimulation due to the high level of IPP they accumulate (Cano et al., 2021); THP-1 (acute myeloid leukaemia cells) are used as a less sensitive target (Witte et al., 2018), and Jurkat (acute T cell leukaemia cells) as non-sensitive targets to Vδ2 T cells (Haecker & Wagner, 1994) The target cells were added at a 1:1 ratio on Day 4 (1 day after OKT3). The same number of target cells from Day 4 were added on days 5 and 6 for Vδ2 cells re-stimulation. On Day 7, all cells at the end of co-culture were collected and stained to obtain expression of PD-1 and LAG-3 and counts of Vδ2 T cells (Figure 4.12 and Figure 4.13).

In Figure 4.12, PD-1 levels are shown to increase from Day 0 to Day 3 in the single- and multiple-stimulated conditions. However, it decreases by Day 4 before re-increasing on Day 7 following single, dual, or triple addition of target cells. LAG-3 expression patterns are similar for all three target cells (Daudi, THP-1 or Jurkat). LAG-3 levels follow similar patterns from Day 0 to Day 7, where it increases slightly on Day 3 and then decreases on Day 4, then is upregulated again on Day 7 but is higher in the single-stimulated (ZOL) Vδ2 cell than in the double-

stimulated (ZOL+OKT3) Vδ2 cells. No statistical significance was observed. Figure 4.13 shows a representative donor histogram of PD-1 and LAG-3 expression between single (ZOL) and multiple stimulations (ZOL+OKT3).

Cell counts of Vδ2 T cells (Figure 4.14) showed no apparent differences between single or double-stimulated cells when co-cultured with different targets. Though there are higher Vδ2 T cell counts with Jurkat re-stimulation for following both the single and double stimulated samples, especially with three times Jurkat stimulated Vδ2 T cell where there is an apparent increase in the cell counts for the single stimulated cells compared to the double stimulated cells.

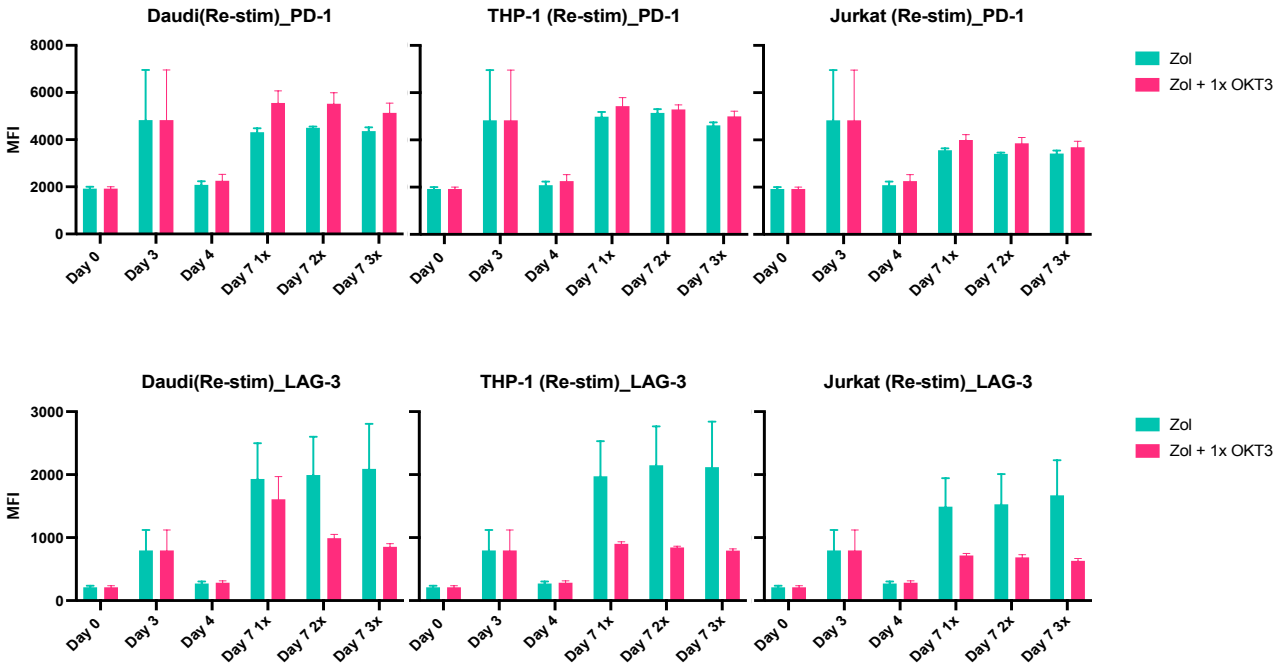



*Figure 4.12: Mean fluorescence intensity of PD-1 and LAG-3 expressed on $V\delta2$ T cells in response to different targets (Daudi, Jurkat or THP-1). Data is shown for single (ZOL) and double (ZOL + 1x OKT3) stimulated $V\delta2$ Cells. Triplicates of Three biological donors were used. No significance is shown when comparing Zol and Zol + 1xOKT3.*

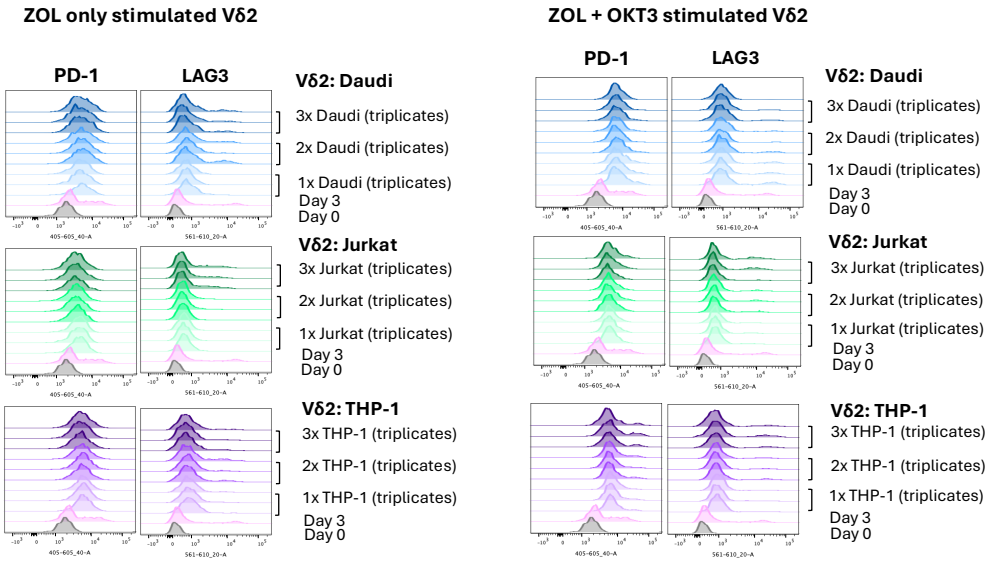



*Figure 4.13: Histograms of PD-1 and Lag-3 in single (ZOL) and double (ZOL+OKT3) stimulated $V\delta2$ T cells in response to 1, 2 or 3 stimulations by different target cells (Daudi, Jurkat or THP-1). Triplicates of three biological donors were used. One representative donor is shown.*

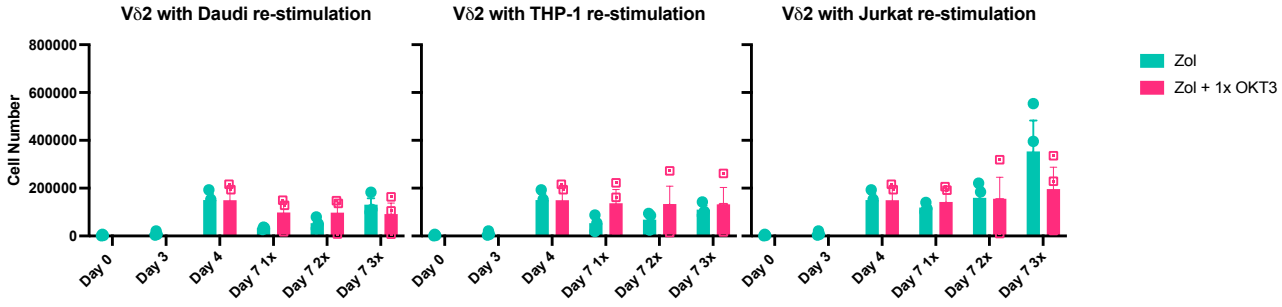



*Figure 4.14: Cell counts of $V\delta2$ T cells in response to different targets (Daudi, Jurkat or THP-1). Data is shown for single (ZOL) and double (ZOL + 1x OKT3) stimulated $V\delta2$ Cells. No significance is shown when comparing Zol and Zol + 1xOKT3. Triplicates of three biological donors were used.*

In summary, the re-stimulation model was developed to test whether early re-stimulation with target cells (24, 48 and 72 hours after OKT3 stimulation) will result in hypoproliferation and exhaustion. The results do not demonstrate hypoproliferation with double stimulation, and the difference in exhaustion markers expression (LAG-3 and PD-1) is non-significant. Therefore, the

effects of tumour cell targeting following initial ZOL and OKT3 stimulation do not show clear evidence of functional exhaustion in Vδ2 T cells.

## Conclusion

Functional experiments were designed to assess the differences in functionality and exhaustion markers expression of single and double-stimulated cells in response to stimulating with different target cells. The first experiment evaluated the proliferation and survival of Vδ2 T cells when co-cultured with Daudi target cells at different effector-to-target ratios. The second experiment was designed to study the emergence of hypo-functionality and heterogeneity of Vδ2 T cell populations following the dual stimulation with OKT3. The final experiment assessed the response of Vδ2 T cells following double stimulations by cell counts and expression of exhaustion markers (PD-1 and LAG-3) in response to re-challenge with different targets with different levels of stimulating potential to Vδ2 T cells (i.e. Daudi, THP-1 and Jurkat) on alternating days.

This chapter evaluated the varied responses of single and dual-stimulated Vδ2 T cells in response to various cancer targets. First, we wanted to assess the proliferation and survival of Vδ2 T cells against Daudi cells. Interesting effects were noted of Daudi cells on reducing Vδ2 numbers without impacting their proliferation. However, no significant effects were shown concerning the different responses of single-stimulated Vδ2 cells compared to the double-stimulated. More informative effects from this experiment could be achieved by expanding the timeline for Daudi cells to efficiently stimulate the Vδ2 cells in culture. Collecting samples for analysis after 72 hours of co-culture might have resulted in differences in responses of the ZOL only and ZOL/ OKT3 stimulated Vδ2 cells. The lack of proliferation of Daudi cells is unexpected and might indicate an issue with the cells grown and used in this study, or to a previously unreported effect of the relatively low numbers of Vδ2 used in the co-culture.

Second, the proliferation assay was set to measure hypo-functionality in the double stimulated of Vδ2 compared to single stimulated. Again, this experiment failed to show significant differences between the different conditions. A primary reason is the complexity of capturing dynamic

changes in proliferation while only looking at snapshots at set time points. Briefly, the cell heterogeneity and preferential stimulation could impact how the proliferation plots look over time. Cells lost due to activation-induced cell death are also not detected with this assay. In contrast, terminally differentiated cells, like terminal effector T cells, would be complex to determine as they will remain non-proliferating at the time points tested. Nevertheless, this does not negate the fact that an exhausted Vδ2 population is present but is much harder to detect due to the heterogeneity of cultures. The use of additional markers of proliferation like Ki67 or including more time-points for short-time detection of proliferation every 12 hours might have shown more differences in proliferative potential. Moreover, expanding the detection points to day 10 or 13 might have shown more informative responsiveness.

Finally, the re-stimulation model was developed to test whether early re-stimulation with target cells (24, 48 and 72 hours after OKT3 stimulation) will result in hypo proliferation and exhaustion. The results do not show hypo proliferation is observed with double stimulation, and the difference in exhaustion markers expression (LAG-3 and PD-1) is insignificant. This specific experiment was analysed at a single time-point after the addition of targets to stimulate the Vδ2 cells in culture. The study did not show efficient variations with the different targets used. A potential limitation of this study was the re-use of a similar number of cells added at the first time point of stimulation; the reason behind this was the inability to adequately calculate the numbers of target cells to Vδ2 cells at the second or third stimulation time-points. This might have caused the predominance of target cells in culture or difficulty in assessing conclusions without taking into account changes to cell numbers at the time of the second or third stimulation with targets.

In summary, early stimulation with target cells after OKT3 addition or studying hypo proliferation in a short-term study is insufficient to convincingly conclude that an exhaustion phenotype is induced between days 4 and 6 in the ZOL + OKT3 stimulation model of Vδ2 T cells.

# Chapter 5

## RNA Transcriptomics of Vγ9Vδ2 T cells

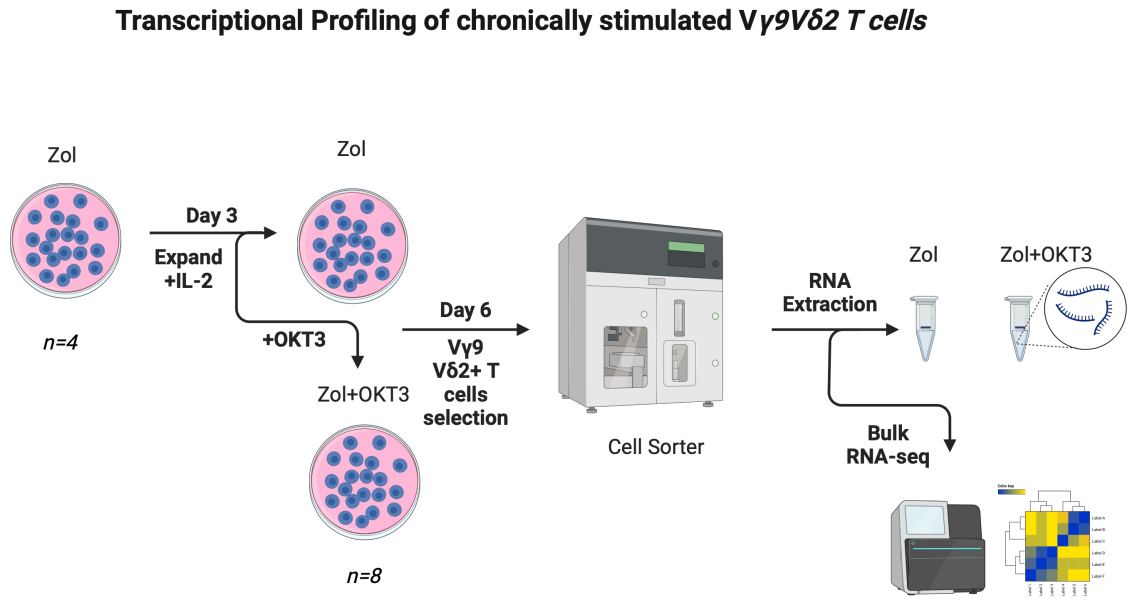
### Introduction

RNA transcriptomics is crucial for understanding the differences between cell states regarding their transcriptomic landscape. RNA seq has been a valuable tool for many researchers to confirm or discover the association or contribution of certain proteins in specific cells and its effects on cell development. For this reason, bulk RNA seq was used to investigate the difference between chronically activated γδT cells and Zoledronate-only expanded γδT cells. Studying the transcriptome will show more profound variations regarding the expression of specific transcription factors or proteins between the two cell types. Such an experiment will further enhance our views on critical stimulation effects and exhaustion development in γδ T cells. It will also shed some light on the similarities or differences between γδ T cells and their conventional counterparts, alpha-beta (αβ) T cells.

Exhaustion has been widely studied in alpha beta (αβ) T cells for the past few years. It has been identified as a crucial research priority for understanding how to make better therapeutics that persist in chronic diseases like cancers. We hypothesise that a state of exhaustion and cellular dysfunction exists in γδ T cells and can be identified by specific markers that are common or distinct from exhaustion markers expressed in conventional αβ T cells. However, as this has not been studied yet in γδ T cells, it would be of great value to understand the transcriptome of chronically activated γδ T cells to develop new therapies and engineering strategies to overcome their exhaustion or lack of persistence.

The experimental set-up prior to RNA profiling is shown in schematic 5.1. Two RNA sequencing experiments were done. Initially a small scale experiment was performed for 3 donors with relaitvely low collection of cells (around 0.2-1 million cells) for RNA processing. A second experiment was then performed for 4 donors and with maximum collection of cells (more than 1 million cells/Donor) and with the aim to increase RNA yield and quality and increasing

the total number of donors tested for exhaustion RNA signatures. Bioinformatics analysis was performed independently for each experiment as they were handled differently by different facilities and using different sequencing devices. It was important to handle the analysis individually to minimize batch effects and variations that could impact the quality of analysis and results interpretation.




*Schematic 5.1: Experimental set-up of Vγ9Vδ2 T cells RNA profiling. Four biological donors are used for each condition after stimulating with Zol on Day 0. On Day 3, the donors are split into two groups (Zol only and Zol + OKT3 stimulated samples). On Day 6, the samples are sorted for positive selection of Vγ9Vδ2 T cells, followed by RNA extraction and Bulk RNA Sequencing.*

## Experiment 1: RNA profiling of Vδ2 T cells

### *Principal component analysis: PCA*

PCA (principal component analysis) is a statistical method for data dimensionality reduction while retaining sample variations. It shows how different conditions cluster and the variation or

similarity between the conditions studied. The distance between points in PCA plots reflects the degree of difference in gene expression profile between the samples, where short distances indicate similarity and long distances are proportionate to the degree of difference between samples. PCA has as many principal components as variables in the data set. The first PC (PC1) explains the maximum variation between variables; thus, it holds the maximum amount of information. The second PC (PC2) explains the second highest degree of variations and information after PC1. PC3 explains the third highest degree of variations and information between the variables.

In the first (small scale) experiment, the PCA plots shown (Figure 5.1) are for PC1 and PC2 (in A) and PC1 and PC3 (in B), meaning the highest variability is demonstrated between the samples studied.

In Figure 5.1A (PC1 and PC2 plot), there is a considerable variation between all samples regardless of condition. Control samples stimulated with Zoledronate only are shown as ZOL in the analysis, whereas double-stimulated samples with Zoledronate and OKT3 stimulation are shown as OKT3. The initial observation is that there is broad clustering of the two different conditions (single and multiple stimulated) indicating real differences in their transcriptional signatures. From the four control donors, D1 and D2 seem to show similarity and D4 and D3 are also shown to be similar. However, there is also a similarity between some double-stimulated samples (OKT3) and the control samples, as shown by D2 OKT3 and D1 ZOL. B has more apparent clustering and separation between 2 double-stimulated samples and all four controls (single stimulation. However, D4OKT3 is still showing some similarity to some of the control samples and not the treated conditions. D1ZOL and D2ZOL control samples continue showing more similarity than with D3ZOL and D4ZOL, and the same is observed regarding D3ZOL and D4ZOL versus D1ZOL and D2ZOL control samples.

Variations are always expected when using biological donors, especially with γδ T-cells. Increasing the donor pool and cell number might be required to precisely identify any biological variations between the donors and not only transcriptional changes that result from the second stimulation. However, as suggested by PCA, samples show some clustering according to

treatment condition, suggesting changes in transcriptional signatures following double stimulation rather than donor variability. Hence, it is interesting to investigate the genes associated with different stimulatory conditions.

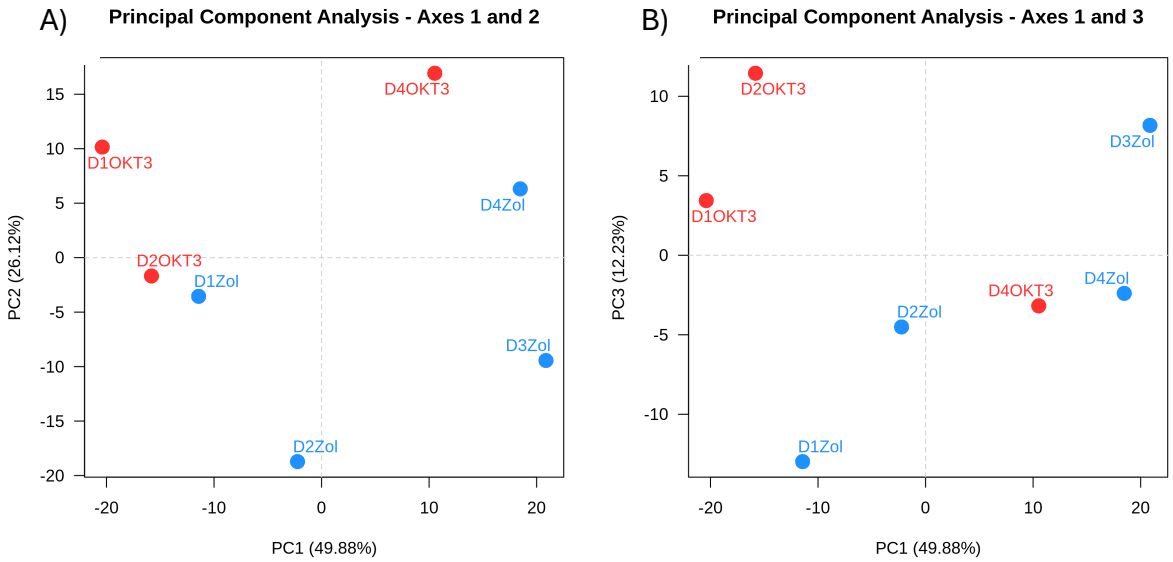



*Figure 5.1: Principal component analysis of both groups studied (control (ZOL), n=4), represented as blue dots and double stimulated (OKT3, n=3), represented as red dots. On the left, PC1 and PC2 are used to make the plot. On the right, PC1 and PC3 are used to make the plot.*

A volcano plot of both groups tested is shown in Figure 5.2. In Volcano plots, the fold change of the gene expression in both sets is plotted against the significance of that change to show how different the expression of all genes in these samples is. The dots in Red are for genes that are significant for being differentially expressed in the double-stimulated samples compared to the control (ZOLEDRONATE (ZOL)). All values in red are of P value less than or equal to 0.05. The right side (positive x-axis) shows the log fold change for upregulated genes, while the points on the left are for downregulated genes. Significantly upregulated genes were 202 in number, whilst there were 232 significantly downregulated genes. Further functional analysis will be shown for the genes identified.

An adjusted p-value (also known as False discovery rate (FDR)) is used for the analysis presented. In RNA-seq, the p-adjusted value is a recent approach used when making multiple comparisons, which controls the number of false discoveries and results in fewer false positives than non-adjusted p-values, thus increasing the reliability of the results (Li et al., 2012).

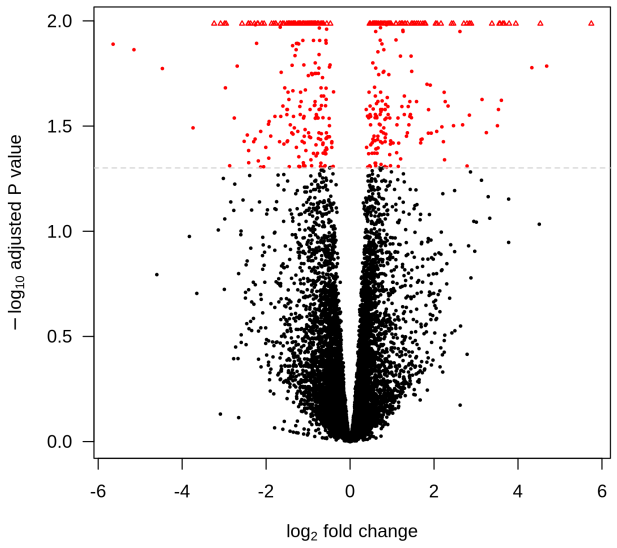



*Figure 5.2: A volcano plot of differentially expressed genes in both griuped studied (Control (ZOL) and double stimulated (OKT3). The results show are for differences in OKT compared to ZOL. The red dots are for significant genes with an adjusted pvalue => 0.05.*

## *Hierarchical clustering based on gene expression*

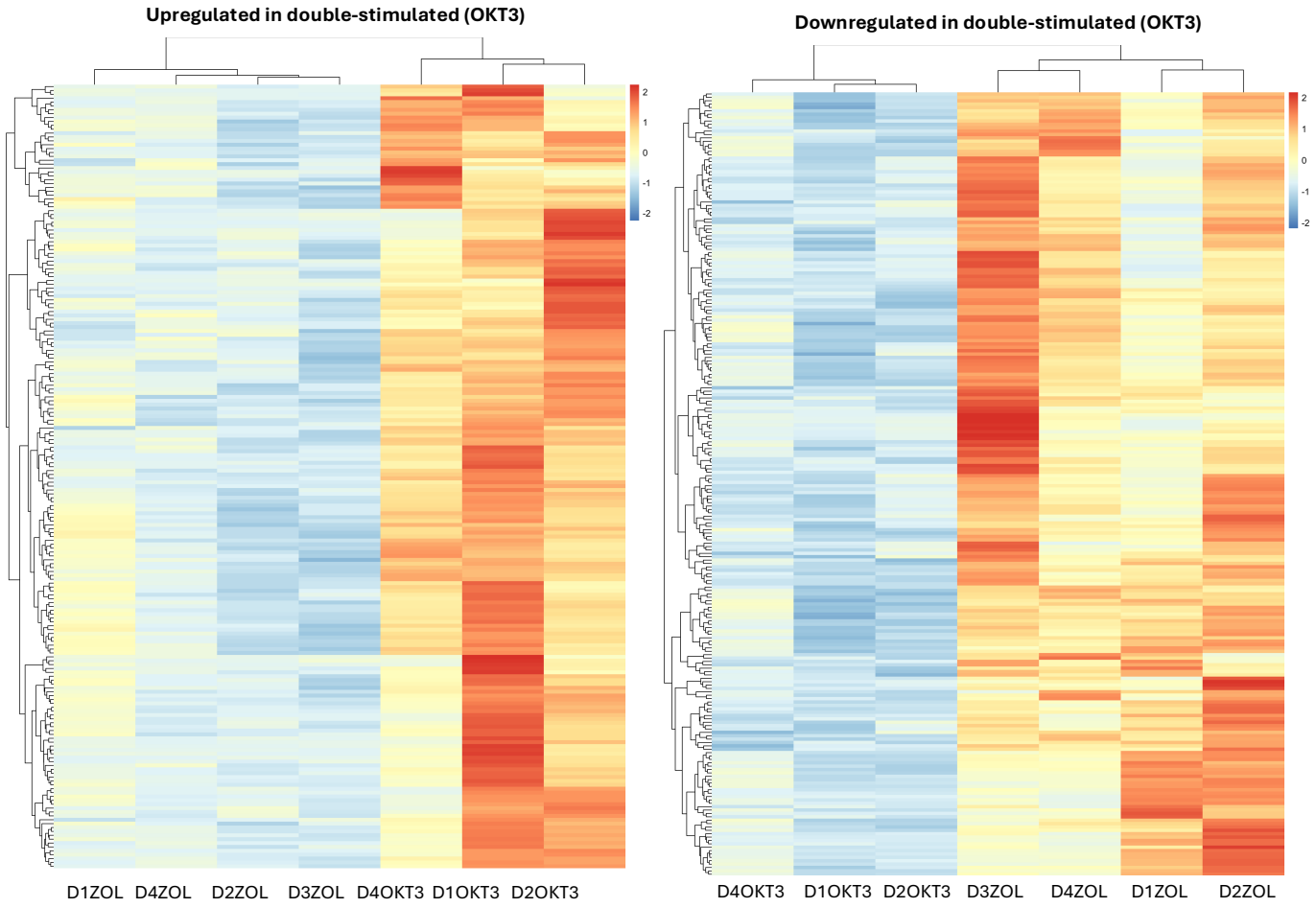



*Figure 5.3: A global heatmap for gene expression in the control (ZOL) samples and double stimulated (OKT3) samples. The full list of transcriptomes is shown here. On the left, the upregulated genes in double stimulated and downregulated in the control group are shown. On the right, a heatmap of genes downregulated in the double-stimulated samples and upregulated in the Control are shown.*

Heatmaps are commonly used to normalise gene expression data. This analysis combines the level of gene expression per sample per gene, with clustering to show similarities between samples in terms of their gene expression patterns. Top clustering shows similarity between samples, while side clustering shows similarity between genes. The colours show changes and

variations in gene expression between samples from low (in blue) to high (in red); Genes in which there is no change between conditions are shown in yellow.

A heatmap is shown for the genes expressed in the samples (Figure 5.3). It is demonstrated that Zoledronate samples (control or ZOL) cluster together and separately from double-stimulated samples (OKT3). As can be seen, the upregulated genes are highlighted here for the double-stimulated samples (OKT3), and gene expression patterns between all the samples can be seen. For OKT3 samples, the upregulated genes are higher in number for one of the donors (Donor 1) in the middle compared to the other two donors. Every sample has its distinct pattern. When we look at the Control samples (ZOLEDRONATE (ZOL)), D1 on the far left seems to have lower expression pattern variations than the other three donors shown. It is also interesting that previously (in Figure 5.1A), this donor was clustering with double stimulated samples, and in Figure 5.1B, using PC1 and PC3, D1 was an outlier. This data suggests D1 gene expression pattern following ZOL stimulation was more akin to double stimulation or that D1 transcriptional response to OKT3 was more marked that in the other donors. The second heatmap shows the opposite, while the downregulated genes are highlighted in the double-stimulated samples, and upregulated genes are shown for the control group. The heatmap on the right shows that D3 and D2 of the control have a slightly different expression pattern from the two other donors tested. The upregulated genes are higher in D3 ZOL and D2 ZOL.

The hierarchical clustering showed different transcriptional patterns in the double-stimulated samples (OKT3) compared to the control. Clustering is consistent with what was shown in PCA in Figure 5.1. Interestingly, despite differences regarding biological variations between donors, distinct differences are exclusively due to differential stimulation of the populations. Overall, the finding of consistent clustering of gene expression patterns indicates there are significant and consistent patterns in the respective ZOL and OKT3 conditions, and hence that gene lists of the most significantly different gene between the two conditions might be reflecting of true and generalizable transcriptional differences following the second stimulation.

*Differentially expressed genes: Upregulated*

A box plot of the top 25 upregulated genes in the OKT3 conditions, and their molecular interactions are shown (Figure 5.4 A and B). Of the genes selected, only BATF3 has been described before as associated with T-cell exhaustion. BATF3 is from the exhaustion related AP-1 family of transcription factors and is associated with JUN by heterodimerisations and binding to DNA to regulate the expression of target genes. Increased expression of BATF3 has been demonstrated to trigger the upregulation of inhibitory molecules in T cells and dysregulation of activating AP-1 family of transcription factors (c-JUN/c-FOS), leading to exhaustion in αβ T cells (shown in network 1)(Seo et al., 2021). Increased expression of BATF3 in the samples following putative exhaustion stimulus (double-stimulations) is supportive that exhaustion is initiated in the multiple stimulated samples.

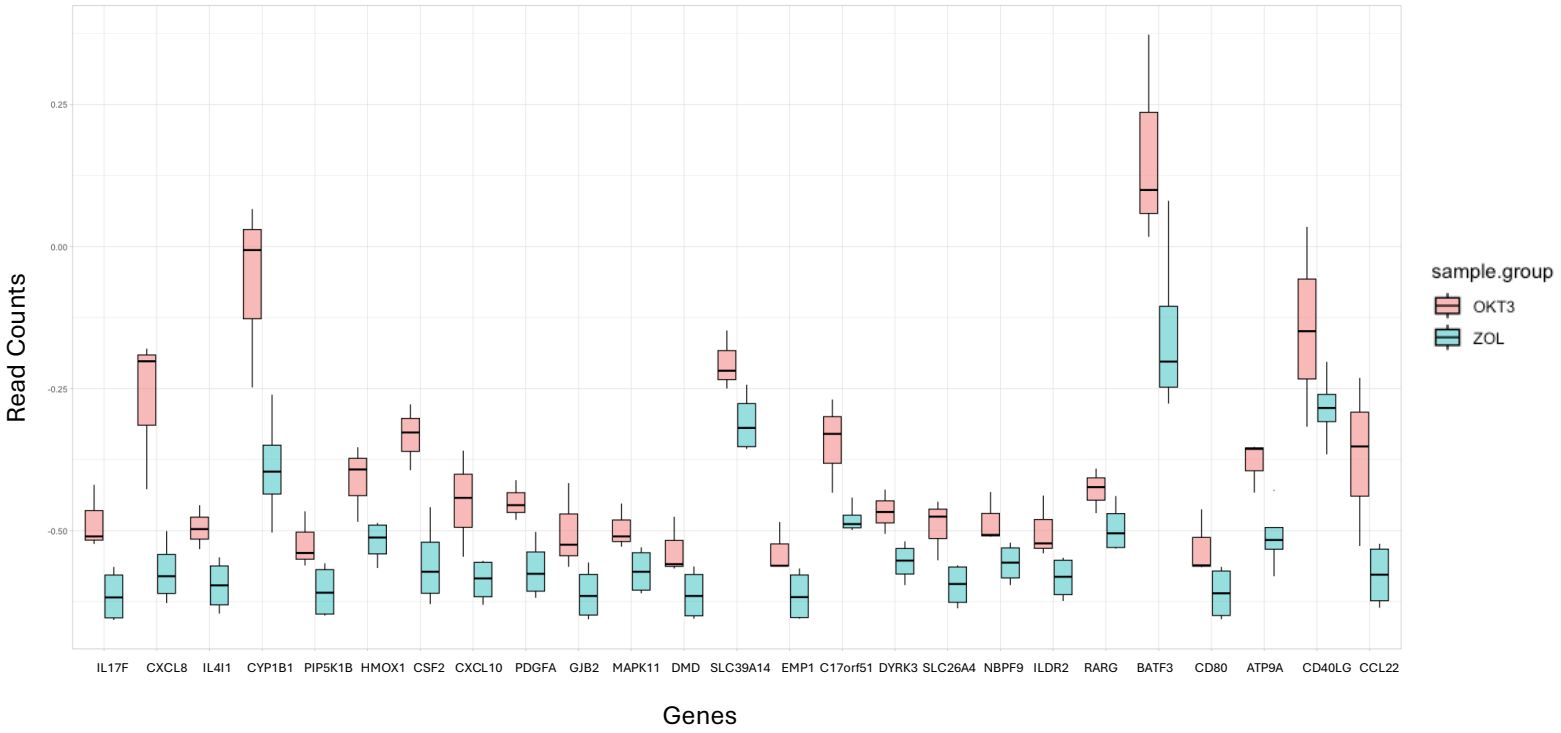
STRING is a database of known protein-protein interactions, and it shows known and predicted interactions that are direct or indirect to the input query (the list of differentially regulated genes). For STRING analysis, one or multiple gene names can be investigated for their potential interactions with other proteins, providing more insights into the potential roles of selected proteins. For the following analysis, significantly differentially expressed genes were used as input for STRING protein query software to look up potential networks of exhaustion-associated genes within the interactions of each protein. Two networks showed interactions with known exhaustion genes and are shown in Figure 5.4B. STRING analysis revealed that as well as BATF3, HMOX1 also lies within a network of exhaustion-related genes that involves JUN and FOS, the canonical components of the AP-1 transcription factor complex which is known to promote survival of T cells (shown in network 2) (Funes et al., 2020).

The following, from the list of differential expressed genes, have shown evidence of their contribution to exhaustion-related phenomena in the literature. IL17F (Interleukin 17F ), the highest-ranked gene on the list in terms of log fold change, is an inflammatory cytokine known to have a role in immunity against pathogens in IL17-producing cells like CD4+ T cells and gamma delta T cells; it also stimulates other cytokines such as CSF2 (colony-stimulating factor 2 ) (also upregulated in the list)(Chang & Dong, 2009). PDGFA (platelet-derived growth factor

receptor alpha) is a growth factor that promotes cell proliferation and survival in T-cell lymphoma (Piccaluga et al., 2014). MAPK11(Mitogen-activated protein kinase 11) is a part of the MAP kinase pathway and is crucial in responding to external stimuli such as proinflammatory cytokines or physical stress. It has been found to have a role in multiple cancers, but it is still not fully understood in the context of gamma delta T cells (Roche et al., 2020). CD80 (Cluster of differentiation 80) is an activation antigen or, more specifically, a costimulatory ligand expressed on antigen-presenting cells, which interacts with CD28 (T-cell-specific surface glycoprotein CD28) receptors in T cells. CD80 has also been shown to have inhibitory functions to PD-1 via PD-L1 (Programmed death-ligand 1) and CTLA-4 (Soskic et al., 2021; Sugiura et al., 2022) in line with what is described in the literature about the role of gamma delta T cells as antigen-presenting cells that display costimulatory signals and chemokine receptors upon their activation (Brandes et al., 2005; Himoudi et al., 2012). CD40LG is the gene expressing CD40L (CD40 ligand (type II transmembrane glycoprotein)) on activated T cells. It is necessary for tumour rejection by CD8+ T cells and is also described as a way for CD8+ T cells to promote their growth and proliferation (Marigo et al., 2016; Tay et al., 2017). CCL22 (C-C motif chemokine 22) is a chemokine ligand that interacts with chemokine receptors on the surface of target cells for trafficking to inflammatory regions (Lança et al., 2013).

Although there are genes shown that are exhaustion associated and are shown within a string protein network of reported exhaustion genes (Figure 5.4), the majority of upregulated genes are associated with T cell activation and suggest that OKT3 stimulation resulted in an activation signature of T cells and no evidence of a strong initiation of exhaustion signature can be concluded.

(A)




(B)

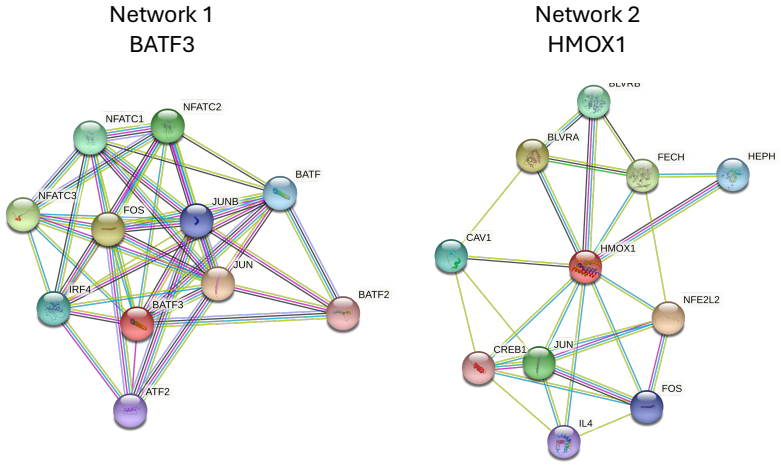



*Figure 5.4: The top 25 significantly upregulated genes (ranked left to right based on their significance) in the double stimulated sample are shown in (A). Networks from string protein analysis with potential interactions of selected genes (BATF3 and HMOX1) with other proteins are shown in (B).*

### *Differentially expressed genes: Downregulated*

A boxplot of the top 25 downregulated genes in the OKT3 (double-stimulated) sample group compared to the Zoledronate group (ZOL) is shown in Figure 5.5. From the 25 genes shown (Figure 5.5), three genes have been described previously in the literature with an association with T cell development and behaviour. MXI1 (MAX interactor-1) is a transcriptional repressor antagonising MYC (proto-oncogene) activation, which leads to growth-related gene activation (Lei et al., 2022). WNT10B (Wnt Family Member 10B) is part of the WNT ligand gene family, and it activates Wnt signalling, leading to the control of stemness and cell fate (Perkins et al., 2023). NDRG1 (N-myc downstream-regulated gene) is another protein that is found to be associated with tolerance in B cells and is an anergy factor for CD4+ T cells. However, its role has been considered unnecessary for T cells anergy (Hodgson et al., 2022). Twenty-two genes in the list are not previously been reported to directly influence specific immunoregulatory pathways or activation/repression of pathways, and it is not fully understood what their roles are in cell signalling or survival.

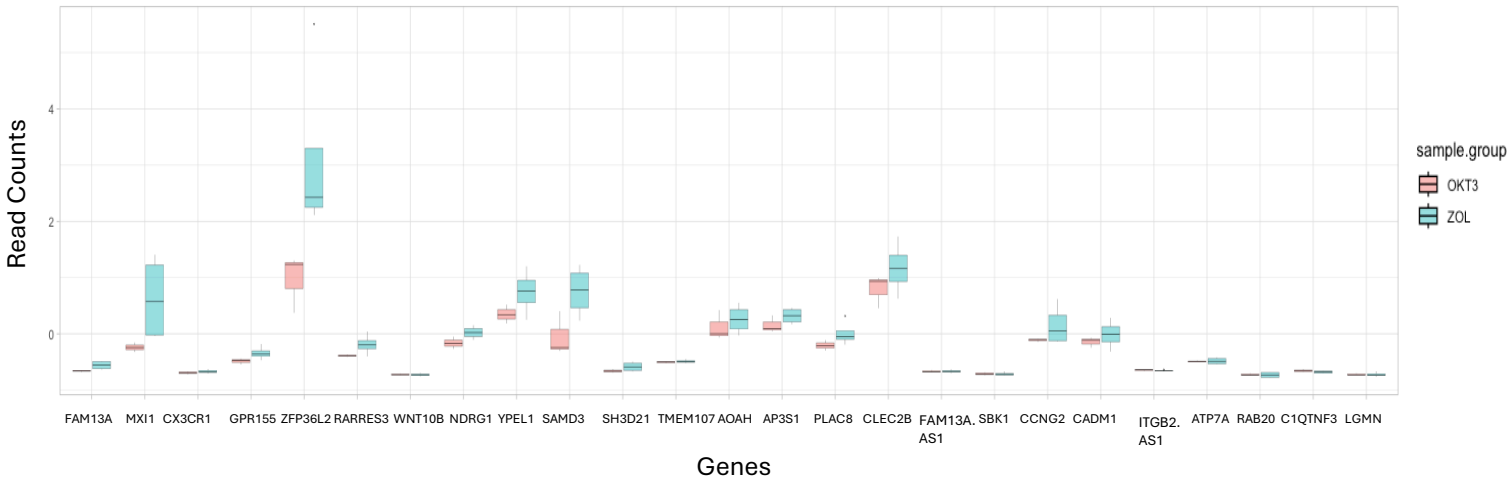



*Figure 5.5: The top 25 significantly downregulated genes (ranked left to right based on their significance) in the double stimulated sample are shown.*

*Functional Analysis*

### Over-representation analysis (ORA): Biological Processes

ORA is a method to investigate the potential roles of a list of genes. It finds known pathways with an overrepresentation of genes of interest to give insights into their biological and molecular functions. The biological process ORA specifically assigns genes to a biological process that contains the set of genes enquired. For this data set, the complete list of genes was filtered, and significant genes with p-adjust values of less than or equal to 0.05 and a log2fold change of 2 or higher were analysed The ORA analysis (Figure 5.6 A and B) was only applied to significant genes within a set parameter. In an ORA analysis, the gene ratio is the ratio of input genes to the annotated pathway, gene count is the number of genes that overlap between the query list and the pathway gene list and a p.value (p. adjust) where increased significance means an increased association of the GenGO pathway to the input genes. The retrieved processes involve mainly immune responses and chemokine pathways. As expected, cytokine-mediated pathways and cell development are also among the biological processes found. Gene ontology (Figure 5.6 B) shows the same processes but also provides a relationship between different pathways, which could be of great interest if more variation in the biological sets were provided. The cnet (concept network plot) in Figure 5.7 lists genes contributing to each biological process. A few genes in cnet are already described in the top 25 genes dot plot for upregulated and downregulated genes (like IL17F, CSF2 and CXC3CR1), while other genes were not shown but are in the complete list of significant genes.

The functional analysis did not immediately identify pathways with clear association with cell differentiation and exhaustion or chronic activation. However, these results will be further analysed using the ORA molecular function and GSEA analysis tools.

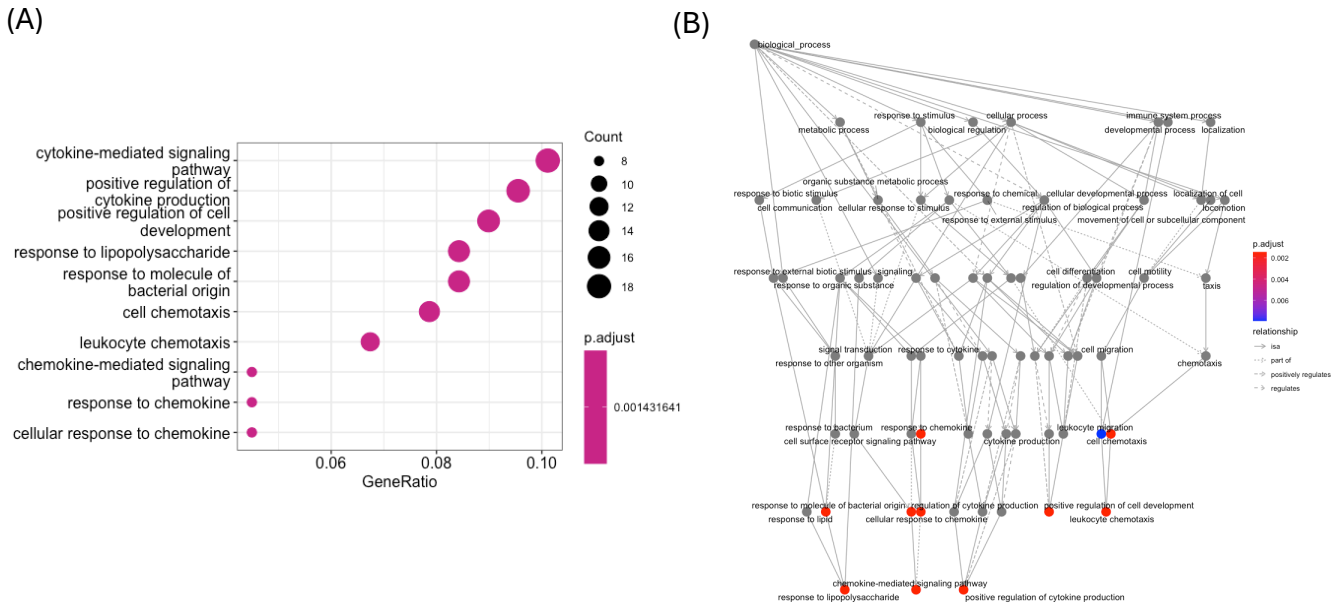



*Figure 5.6: Over-representation analysis in biological process is shown with (A) a dot plot where x-axis is the size of overlap of gene of interest/size of overlap of genesets collection and dots size corresponds to genes counts in a specific set and (B) a gene ontology network of different biological processes sets and their interactions. All data shown are of significant and with p-adjusted value of <= 0.05.*

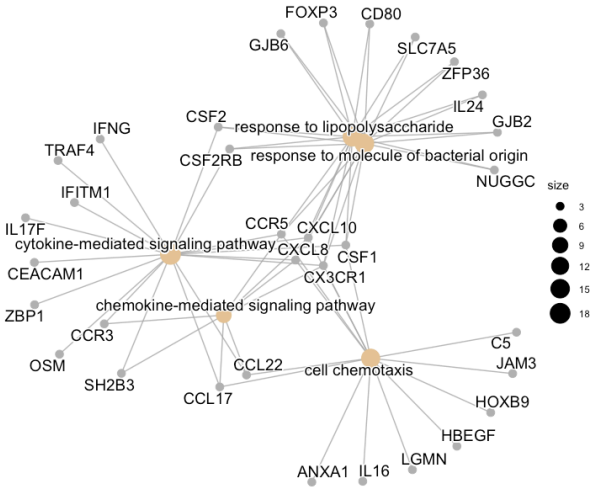



*Figure 5.7: A concept netwrok plot showing all genes from the significant genes list that contributing to over-representation of each of the biological processes sets. Yellow nodes are used for genesets and greay nodes are for genes.*

### Over-representation analysis (ORA): Molecular functions

The molecular functions ORA specifically assign genes to a molecular function set containing the genes of interest. Significant genes with p-adjust values of less than or equal to 0.05 and a log2fold change of 2 or higher were filtered from this data set's complete list of genes. The ORA analysis (Figure 5.8 A and B) was only applied to significant genes within the set parameter. The most extensive molecular function sets are for receptor ligands and signalling, followed by cytokine binding and receptors. The most minor molecular function sets were for chemokine binding and activity. However, only the most significant sets were shown here and sets with p-values higher than 0.01 were not plotted. The GO plot of these 8 significant molecular function associations (Figure 5.8 B) does not establish a complex relationship between these sets as they all have similar functions in general, and few are considered the same rather than with overlapping molecular functions, as shown by the full arrows connecting the nodes. The cnet network in Figure 5.9 shows the specific genes associated with each set and mainly shows genes in the top 25 list, such as IL17F, CSF2, CX3CR1 and Wnt10B.

To understand the involvement of these genes in specific newly published databases, the GSEA will be performed.

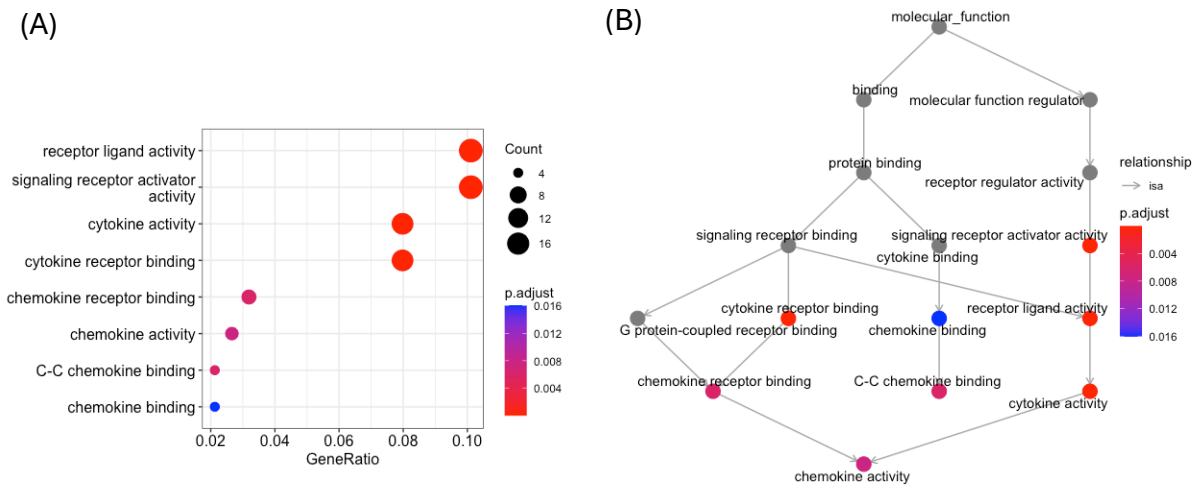



*Figure 5.8: Over-representation analysis in molecular functions sets is shown with (A) a dot plot where x-axis is the size of overlap of gene of interest/size of overlap of genesets collection and dots size corresponds to genes counts in a specific set and (B) a gene ontology network of different molecular functions sets and their interactions. All data shown are of significant and with p-adjusted value of <= 0.05.*

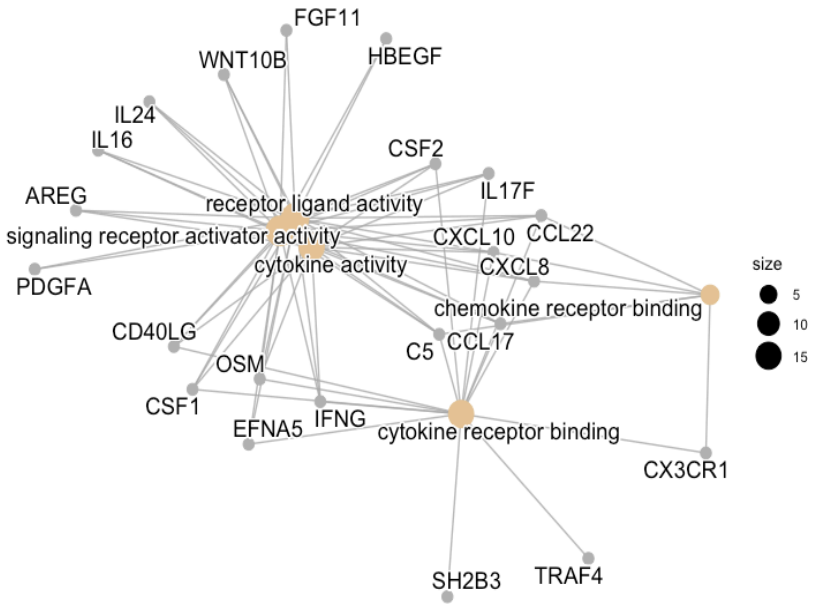



*Figure 5.9: :A concept network plot showing all genes from the significant genes list that are contributing to over-representation of each of the molecular functions sets. Yellow nodes are used for genesets and greay nodes are for genes.*

## Gene set enrichment analysis: Enrichment maps

GSEA is a bioinformatics tool to check for the enrichment of a specific list of genes in published databases. Unlike ORA, where only significantly upregulated or downregulated genes are used as input, in GSEA, the whole list of genes is used to get a fully unbiased view of potential enrichments within a specific database.

### *Hallmarks gene set*

The hallmark gene set contains a list of sets of well-defined biological processes with coherently expressed signatures. It includes 50 groups to screen and will be used as a first choice.

Gene set enrichment analysis was performed on the whole gene set, and sets with a significant p-adjusted value were shown in an enrichment plot. The connecting line between gene sets nodes

indicates an overlap of genes between the two gene sets where the thickness of the line corresponds to the number of genes overlapping. Thicker lines indicate a high number of overlapping genes. The gene set chosen was the hallmarks gene set from the molecular signatures database (msigDB). All the gene sets shown have a significant enrichment from the gene list of the RNA seq data. Table 5.1 shows the set size, NES (normalized enrichment score) and p.adjust for significance. A positive enrichment score corresponds to an enrichment at the top of a pathway's list, and a negative enrichment score signifies an enrichment at the bottom of a pathway's gene list. The order of rank is from most significant to least significant.

Pathways of P53, TNFA signalling via NFKB, IL-2 and STAT 5 signalling are the top ranked enriched lists with genes enriched in the OKT3 versus ZOL-only upregulated genes. This is followed by KRAS signalling upregulated genes, IL-6 JAK-STAT 3 signalling, MYC targets, inflammatory response pathways, UV response upregulated genes pathway, xenobiotic metabolism, and epithelial-mesenchymal transition. The reactive oxygen species pathway has the least degree of positive enrichment and statistical significance that still fulfils the criteria for inclusion.  Other sets are associated with a negative enrichment score and are enriched with genes at the bottom of the genes listed in the pathways shown. These sets are hypoxia, IFN-α response, Glycolysis and IFN-γ response.

The analysis of associations between enriched pathways (Figure 5.10) shows that the enriched pathways with the greatest functional overlap are involved in inflammatory responses and associated with various pathways including KRAS signalling, IL6 and STAT3, TNFA and NFKB, IL-2 and STAT5, and UV response. It also shows genes involved in the p53 pathway and genes involved in processing xenobiotics and other drugs. The epithelial-mesenchymal transition suggests the presence of genes that are involved in wound healing and metastasis. There are less connected sets, such as the reactive oxygen species pathway, which means genes are present that are upregulated in this set. In addition, two MYC target sets are well connected due to similarity and suggest the presence of MYC-regulated genes.

The following sets show significant enrichment with negative NES values, meaning these sets are enriched in the control group (Zol only) instead. Interferon gene sets have the highest

connections between all the groups and indicate IFN-α (Interferon-alpha) and IFN-γ modulation of some genes in the RNA seq set. The glycolysis gene set is connected to the hypoxia gene set, which may suggest similarities. Both sets indicate the presence of genes involved in either encoding proteins for Glycolysis or genes present in hypoxic conditions and related to gamma delta chronic activation.

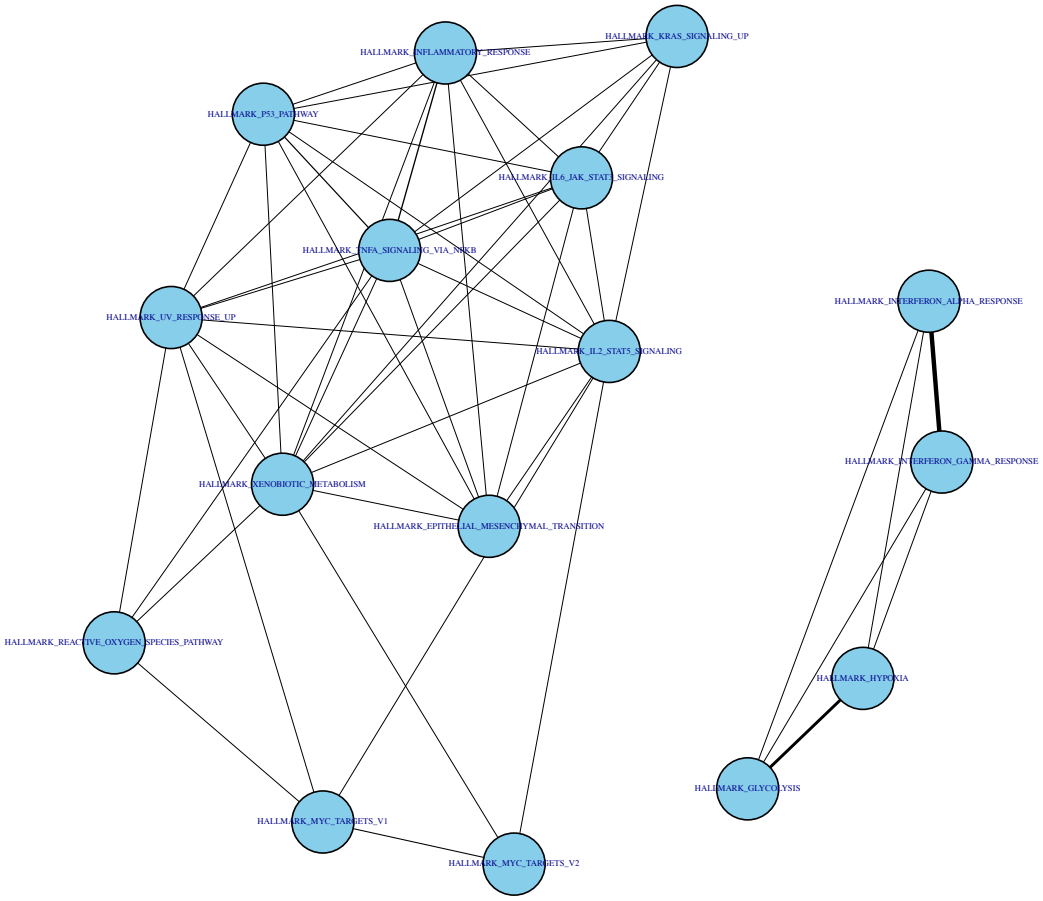



*Figure 5.10: Gene set enrichment analysis of hallmarks genes in the global transcriptome showing significantly enriched sets with p-adjusted <= 0.05.*

*Table 5.1: Enriched hallmark gene sets with set size, normalized enrichment score (NES) and p.adjust values.*

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **Positively Enriched (with positive NES value)** | | | |
| **HALLMARK_P53_PATHWAY** | 164 | 2.01095908 | 3.28426599831944e-06 |
| **HALLMARK_TNFA_SIGNALING_VIA_NFKB** | 163 | 1.97814494 | 5.92830591385055e-06 |
| **HALLMARK_IL2_STAT5_SIGNALING** | 167 | 1.95146928 | 6.14132930762523e-06 |
| **HALLMARK_KRAS_SIGNALING_UP** | 120 | 1.91963919 | 7.98970750483915e-05 |
| **HALLMARK_IL6_JAK_STAT3_SIGNALING** | 59 | 2.01625089 | 0.00022201 |
| **HALLMARK_MYC_TARGETS_V2** | 58 | 1.98739947 | 0.00022201 |
| **HALLMARK_INFLAMMATORY_RESPONSE** | 127 | 1.7563016 | 0.00130345 |
| **HALLMARK_UV_RESPONSE_UP** | 121 | 1.66394104 | 0.00235863 |
| **HALLMARK_XENOBIOTIC_METABOLISM** | 128 | 1.59096469 | 0.00808352 |
| **HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION** | 110 | 1.55000637 | 0.01075584 |
| **HALLMARK_MYC_TARGETS_V1** | 190 | 1.46541776 | 0.01075584 |
| **HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY** | 43 | 1.58790104 | 0.03339249 |
| **Negatively Enriched (with negative NES value)** | | | |
| **HALLMARK_HYPOXIA** | 149 | -2.1639692 | 2.3081555700779e-08 |
| **HALLMARK_INTERFERON_ALPHA_RESPONSE** | 88 | -2.0586583 | 8.49665910311244e-06 |
| **HALLMARK_GLYCOLYSIS** | 157 | -1.4866939 | 0.01243154 |
| **HALLMARK_INTERFERON_GAMMA_RESPONSE** | 170 | -1.4494705 | 0.01583671 |

### *Immune-signatures gene sets (C7)*

Following the hallmarks set enrichment, we next evaluated the immune signatures data set. It is more tailored for immune cells and will provide a better overview of the genes expressed in the double-stimulated samples. It will also highlight similarities between immune cells published using various models to the double-stimulated γδ T cells. Immune signatures can show shared phenotypic traits of other immune cells to the studied OKT3-stimulated γδ T cells.

Many gene sets are shown to have a significant enrichment from the gene list of the RNA seq data, but only the top 20 sets are shown (Figure 5.11 and Table 2) due to the high complexity of the network that is hard to show in a plot clearly. The complete list of significantly enriched

immune-signatures data sets was 772, with 511 sets that are less than or equal to 0.01 p.adjusted value. The gene set size, normalized enrichment scores and p.adjust values are displayed in Table 5.2. The order of rank is from most significant to least significant.

The GSEA analysis from Figure 5.11 shows the identities and inter-relationships of the 20 most significantly enriched gene sets. The top cluster (cluster A) are for genes positively enriched with top genes in each pathway. The most significant enrichment is with the gene list (downregulated) of effector memory CD8 T cells versus naïve CD8 T cells. The second most significant enrichment is with the gene list of downregulated genes of unstimulated vs IL-2 stimulated NK cells and genes downregulated in comparison of untreated CD4 memory cells to 16 hours TSST (toxic shock syndrome toxin) treated young and old donors. Other enrichments include studies of effector and memory virus-specific CD4 T cell subsets in an LCMV model (Hale et al., 2013) and generation of TH1 (type 1 T helper cells) and TH2 (type 2 T helper cells) by CD4 T cells activated with anti-CD3 and anti-CD28 in the presence of TGFB1 (transforming growth factor beta-1) and IL-12 (interleukin-12) for 6 hours versus untreated control at 6hrs (Lund et al., 2003). Other sets are for diseases such as Leishmania and lupus. In particular, the lupus data set means that genes were found that are downregulated in Lupus in CD4 T cells compared to Lupus in B cells. The same was shown for Leishmania as the genes found are upregulated in the uninfected population instead of leishmania-infected dendritic cells after 8 hours.
The bottom cluster (cluster B) in Figure 5.11 is for sets that have a negative enrichment (with genes at the bottom of the pathways), including the same study of untreated CD4 memory cells to 16 hours TSST (toxic shock syndrome toxin) treated young and old donors but for genes that upregulated from the comparison in contrast to downregulated (Hale et al., 2013). Moreover, genes that are downregulated due to low-dose or high-dose of GM-CSF (an immune modulatory molecule) and Curdlan (an agonist of GM-CSF receptor) elicit a robust inflammatory signal to convert immature dendritic cells to potent effectors (Min et al., 2012).

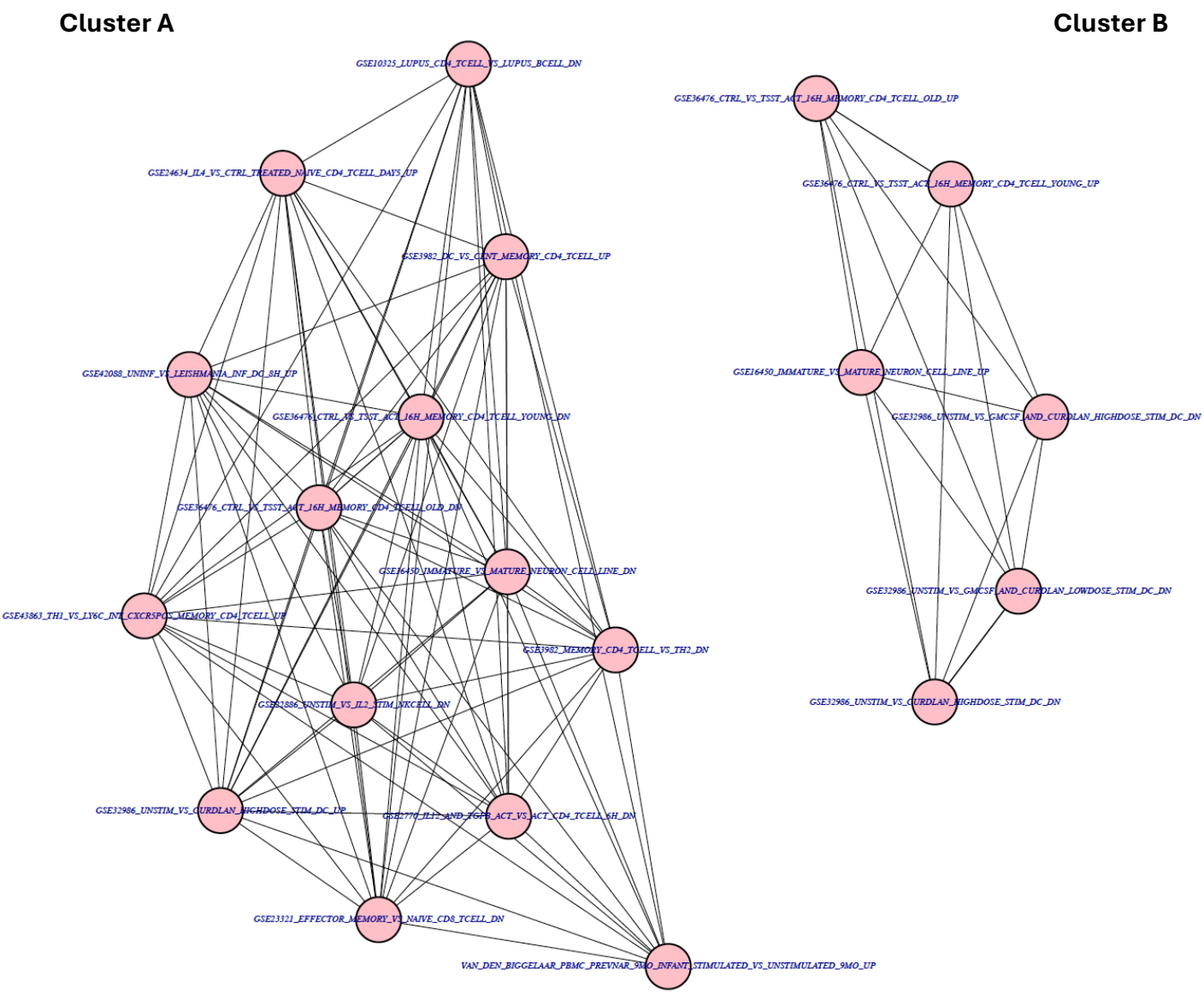



*Figure 5.11: Gene set enrichment analysis of immune signatures genes in the global transcriptome showing significantly enriched sets with p-adjusted <= 0.05.*

Table 5.2: Enriched Immune-signature (c7) gene sets with set size, normalized enrichment score

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **Positively Enriched (with positive NES value)** | | | |
| GSE23321_EFFECTOR_MEMORY_VS_NAIVE_CD8_TCELL_DN | 182 | 2.3589865 | 5.55702419074077e-10 |
| GSE22886_UNSTIM_VS_IL2_STIM_NKCELL_DN | 190 | 2.32579082 | 7.96545179096027e-10 |
| GSE36476_CTRL_VS_TSST_ACT_16H_MEMORY_CD4_TCELL_YOUNG_DN | 182 | 2.32963425 | 1.20213787928108e-09 |
| GSE43863_TH1_VS_LY6C_INT_CXCR5POS_MEMORY_CD4_TCELL_UP | 180 | 2.30504378 | 1.79928598109875e-09 |
| GSE2770_IL12_AND_TGFB_ACT_VS_ACT_CD4_TCELL_6H_DN | 179 | 2.27951155 | 4.43464498769038e-09 |
| GSE16450_IMMATURE_VS_MATURE_NEURON_CELL_LINE_DN | 177 | 2.30508275 | 6.21484987188217e-09 |
| GSE36476_CTRL_VS_TSST_ACT_16H_MEMORY_CD4_TCELL_OLD_DN | 177 | 2.28671621 | 1.18127894894366e-08 |
| VAN_DEN_BIGGELAAR_PBMC_PREVNAR_9MO_INFANT_STIMULATED_VS_UNSTIMULATED_9MO_UP | 94 | 2.48084064 | 1.84872495409182e-08 |
| GSE3982_MEMORY_CD4_TCELL_VS_TH2_DN | 184 | 2.23759696 | 2.11985141352614e-08 |
| GSE42088_UNINF_VS_LEISHMANIA_INF_DC_8H_UP | 142 | 2.33437866 | 3.37448312784837e-08 |
| GSE3982_DC_VS_CENT_MEMORY_CD4_TCELL_UP | 151 | 2.30098793 | 8.28796763490168e-08 |
| GSE32986_UNSTIM_VS_CURDLAN_HIGHDOSE_STIM_DC_UP | 189 | 2.12855212 | 8.73411871424955e-08 |
| GSE24634_IL4_VS_CTRL_TREATED_NAIVE_CD4_TCELL_DAY5_UP | 140 | 2.28040961 | 9.72862869150219e-08 |
| GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_BCELL_DN | 142 | 2.29818296 | 1.10250724739977e-07 |
| **Negatively Enriched (with negative NES value)** | | | |
| GSE32986_UNSTIM_VS_CURDLAN_HIGHDOSE_STIM_DC_DN | 184 | -2.3860961 | 1.09842836198043e-12 |
| GSE16450_IMMATURE_VS_MATURE_NEURON_CELL_LINE_UP | 173 | -2.2454801 | 5.55702419074077e-10 |
| GSE32986_UNSTIM_VS_GMCSF_AND_CURDLAN_LOWDOSE_STIM_DC_DN | 180 | -2.2483492 | 1.32248303738129e-09 |

| | | | |
|---|---|---|---|
| **GSE36476_CTRL_VS_TSST_ACT_16H_MEMORY_CD4_TCELL_YOUNG_UP** | 181 | -2.1889549 | 3.43793207504748e-09 |
| **GSE32986_UNSTIM_VS_GMCSF_AND_CURDLAN_HIGHDOSE_STIM_DC_DN** | 175 | -2.1251931 | 3.60621611353562e-08 |
| **GSE36476_CTRL_VS_TSST_ACT_16H_MEMORY_CD4_TCELL_OLD_UP** | 150 | -2.1683957 | 7.26476632619958e-08 |

## Experiment 2: RNA profiling of Vδ2 T cells

A second experiment was designed to confirm observations with the first bulk-RNA sequencing study. The experiment was aimed to maximise the number of cells collected and increase the RNA yields from the new four donors as one donor from the first time was excluded due to low RNA yield and quality. Moreover, it would be interesting to have a collective view of all the combined analyses, considering different donors might have different transcriptional responses to the same stimulus (ZOLEDRONATE and OKT3). However, as these experiments were done at different times and using different facilities and machines, the analysis will be done separately. Merging the two datasets could introduce batch effects and technical variations that compromise the interpretation of the results and the identification of actual outcomes.

### *PCA*

In RNAseq experiment 2, the PCA plots shown are for PC1 and PC2 and the highest variability is demonstrated between the samples studied.

For this experiment, there is a clear separation between the clustering of Control samples (ZOLEDRONATE (ZOL)) and double-stimulated samples (OKT3) (Figure 5.12). Each dot is for a single donor sample used in this study. PC1 and PC2 are only shown as they represent the highest degrees of separation and are sufficient to provide a clear separation and clustering of the two groups (Zol and OKT3).

Apparent differences between expression profiles are suggested by the clustering of Control samples on one side (Right) and double-stimulated samples on the opposite side (left). The PCA

clustering of different conditions in experiment 2 is more significant compared to experiment 1, implying the technical superiority of this experiment in identifying genes associated with repeat stimulation (OKT3).

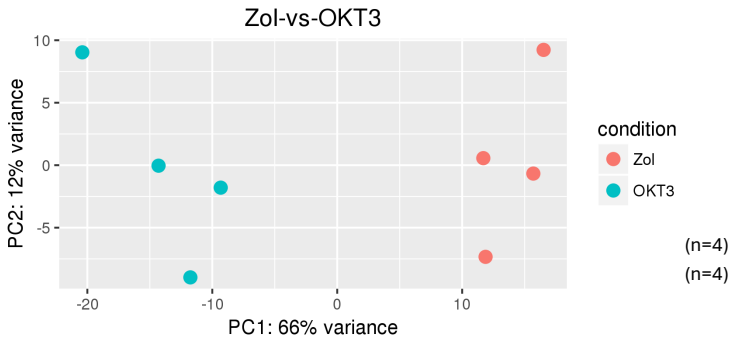



*Figure 5.12:Principal component analysis of both groups studied (control (ZOL) and double stimulated (OKT3). PC1 and PC2 are used here to make the plot.*

A volcano plot of both groups tested is shown in Figure 5.14. The fold change of the gene expression in both sets is plotted against the significance of that change to show how different the expression of all genes in these samples is. The dots in Red are for significantly upregulated genes with a p-value equal to or less than 0.05. The dots in blue are for significantly downregulated genes with a p-value equal to or less than 0.05. The differential expression compares the double-stimulated samples to the control (ZOLEDRONATE (ZOL)). The right side (positive x-axis) shows the points for upregulated genes, while the points on the left are for downregulated genes. As the number of cells used and acquired reads is much larger than in the first experiment, the total number of upregulated genes was 1429. The down-regulated genes were 1953 genes. A comparison between the difference in significant gene numbers is shown in Figure 5.13 and supports that a more thorough conclusion will be made from the second transcriptomic analysis.

Further biological relevance and functional analysis will be shown for the genes identified.

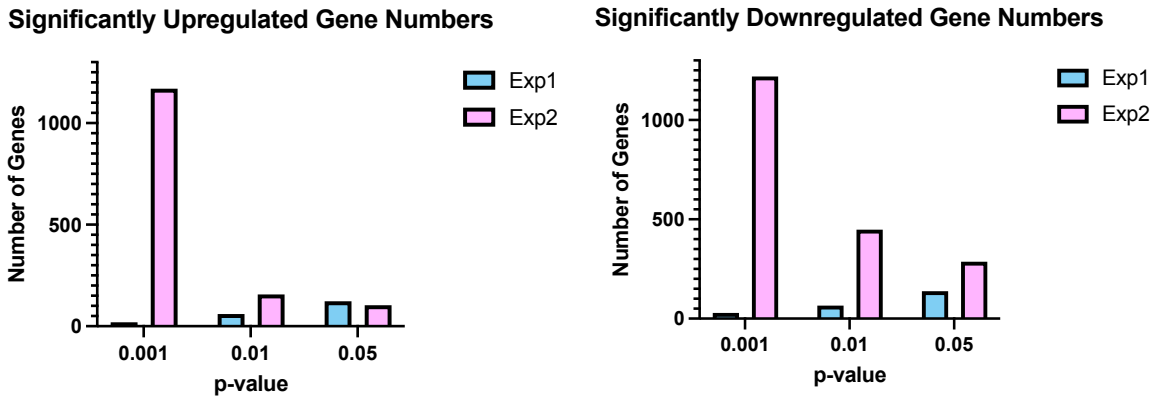



*Figure 5.13: Differentially expressed gene numbers of the two transcriptomic studies, EXP1 (The first study) and EXP2 (the second study). Gene numbers are of significantly expressed genes.*

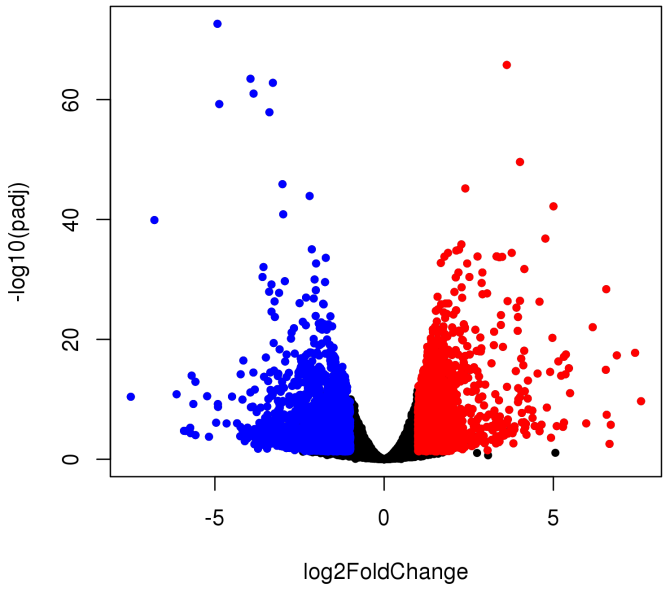



*Figure 5.14: A volcano plot of differentially expressed genes in both griuped studied (Control (ZOL) and double stimulated (OKT3)). The results show are for differences in OKT compared to ZOL. The red dots are for significant genes with an adjusted pvalue => 0.05.*

*Hierarchical clustering based on gene expression*

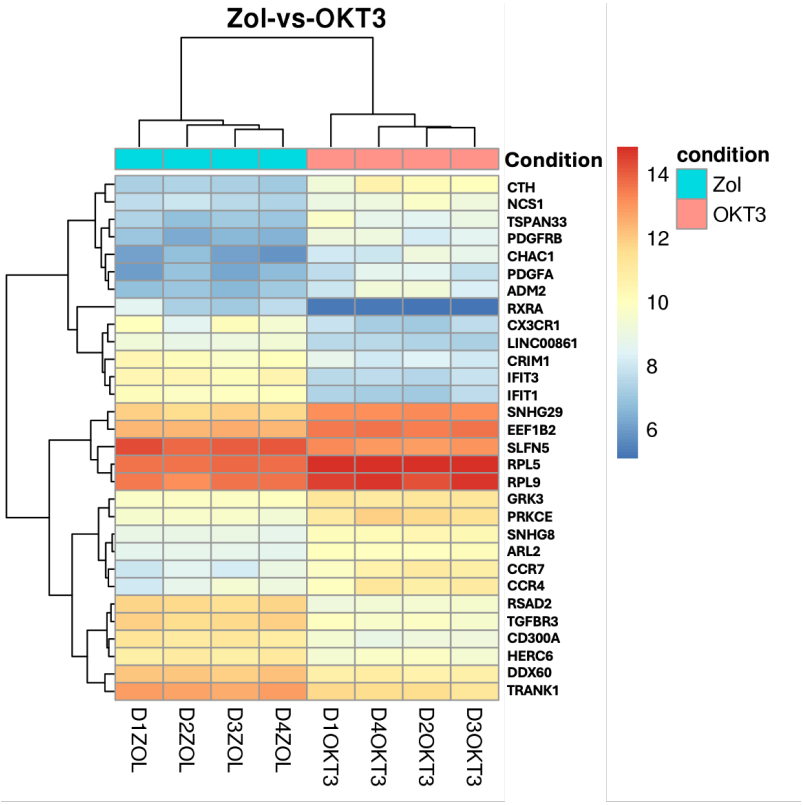



*Figure 5.15: A heatmap of genes with the most significant differential expression in the control (ZOL) and double-stimulated (OKT3) samples. Only significant genes are shown here.*

A heatmap is shown in Figure 5.15 for the top upregulated and downregulated genes from the complete list. The expression scale is from Red (high) to blue (low). In this heatmap, a lot of different patterns can be noticed. Some genes in the top half show opposite trends between the Control (ZOLEDRONATE (ZOL)) and double-stimulated (OKT3) samples. In the middle, the genes are considered upregulated in both groups, but it was highly upregulated in the OKT3 group, especially two genes. The bottom half of the heatmap shows genes that generally have a mid-expression trend and similar patterns between groups. Hierarchal Clustering is also displayed and demonstrates the clustering of Zoledronate samples and OKT3 samples together and agrees with the PCA clustering results shown.

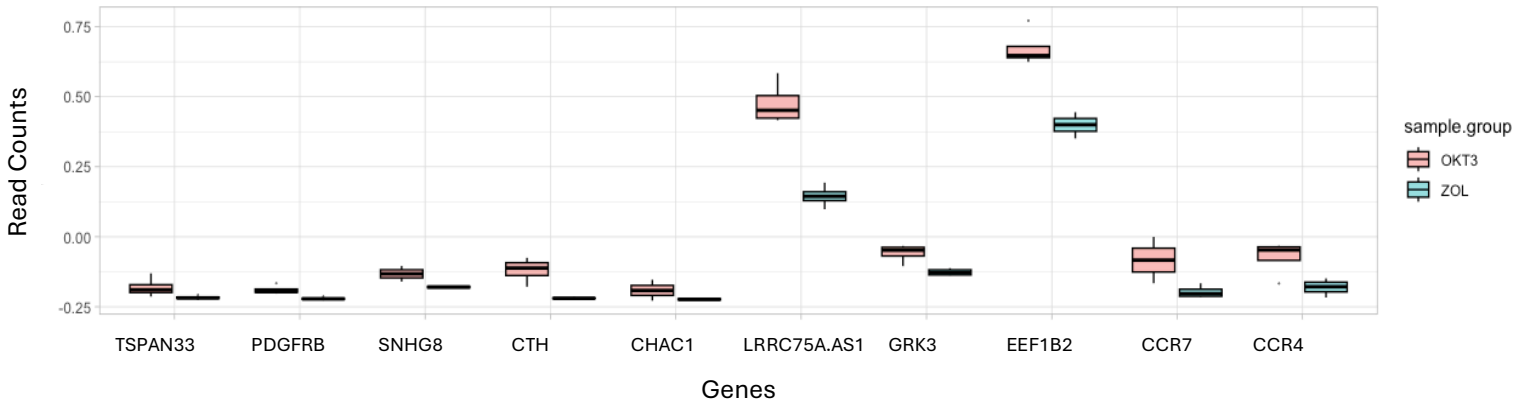
*Differentially expressed genes: upregulated*

From the significantly upregulated ranked gene list, only the top 25 genes are chosen to show differential expression graphically (Figure 5.16). Nevertheless, this experiment had a long list of genes that are many interesting genes that were differentially expressed and upregulated in the double-stimulated samples. A network of top significant genes is shown, and genes associated with known exhaustion-related genes will be described in detail. BATF3, LAG3, TOX2, NR4A, TCF4 (Transcription factor 4), TCF12 (Transcription factor 12) and TCF7L2 (Transcription factor 7 like 2) are genes that have been described in exhaustion in conventional t-cells and are also shown to be significantly expressed in double-stimulated γδ T cells.
LAG3 is an inhibitory receptor associated with exhaustion development (Grebinoski et al., 2022). It is highly expressed on a protein level (from the previous study in Chapter 1, Figure 3.9) and a transcriptional level, as shown from the double-stimulated γδ T cells RNA profiling. TOX2 is associated with the expression of TOX, TOX4, TIGIT, HAVCR2 (also known as TIM-3), TBX21(also known as T-bet) and TCF7, which are known for defining an exhaustion state in T cells (shown in Figure 5.16 C)(Seo et al., 2019). NR4A is correlated with FOS (proto-oncogene) and BCL2 (BCL2 apoptosis regulator) expression is also reported in T-cell dysfunction ((Liu et al., 2019; Seo et al., 2019)). BATF3 was significantly upregulated in both RNA profiling experiments. It is known for promoting t-cell differentiation (Seo et al., 2021).

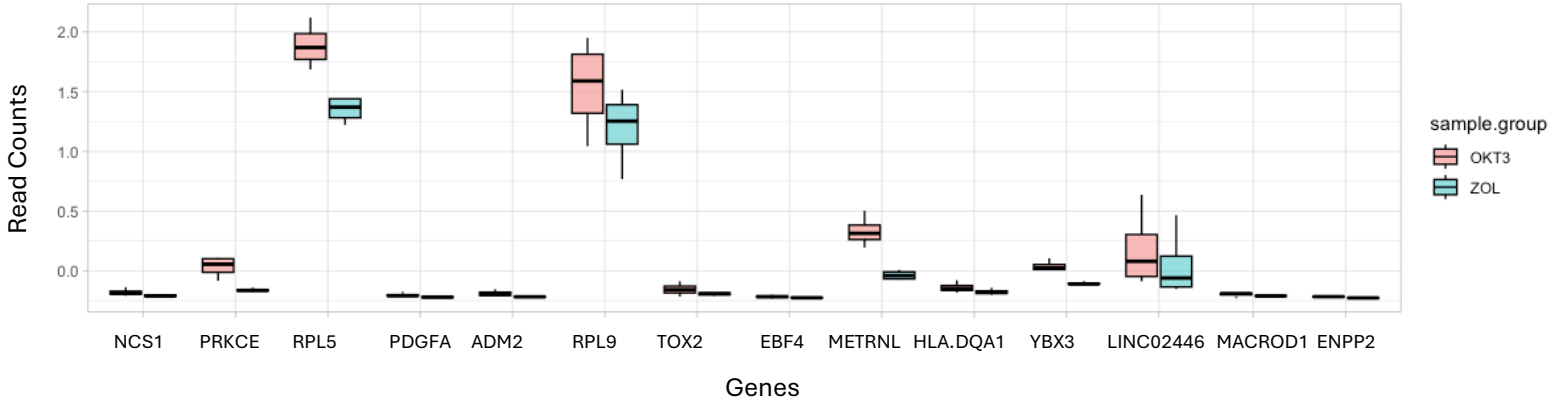
Other genes expressed, like CHAC1 (glutathione specific gamma-glutamylcyclotransferase 1) and PRKCE (Protein kinase C epsilon), are associated with apoptosis, where CHAC1 contributes to apoptosis initiation by glutathione cleavage (Crawford et al., 2015). Some listed genes are part of or mediators to G protein-coupled receptors. PRKCE is an anti-apoptotic gene involved in multiple cellular processes, including cancer survival mechanisms and immune responses(Kumar et al., 2015; Wang et al., 2022). PDGFA (platelet-derived growth factor receptor alpha) was significantly overexpressed in double-stimulated γδ T-cells and is a growth factor promoting cell proliferation and T-cell lymphoma survival (Piccaluga et al., 2014). EBF4 is a transcription factor for the Olf/EBF family of helix-loop-helix transcription factors and has a role in cytotoxic T cells (Kubo et al., 2022)

STRING analysis was performed on TOX2 specifically as it is validated for its role in the context of exhaustion, specifically in T cell development and memory development (Collins et al., 2023; Seo et al., 2019).

(A)




(B)



(C)

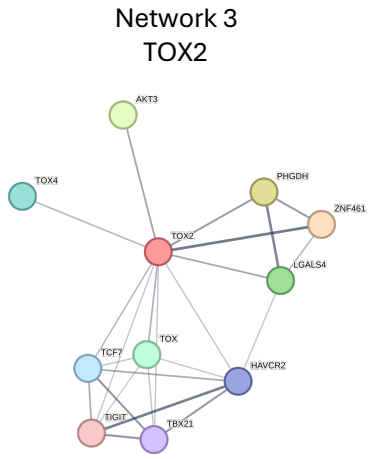


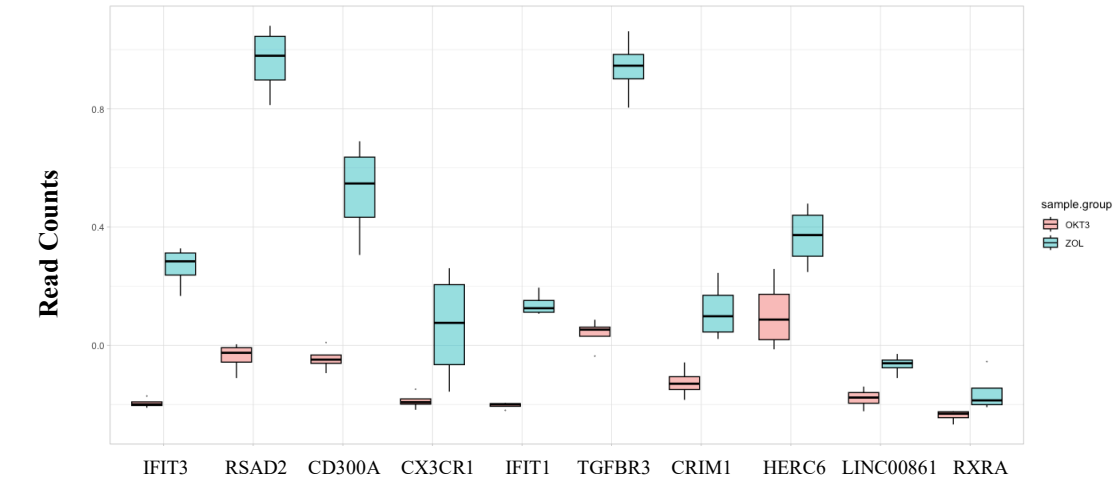

*Figure 5.16: The top 25 significantly upregulated genes in the double stimulated sample are shown in (A) and (B). Networks from string protein analysis with potential interactions of* TOX2 *with other proteins is shown in (C).*

### *Differentially expressed genes: downregulated*

A boxplot of the top 25 downregulated genes is shown (Figure 5.17). The genes are ranked in order of significance from left to right. Out of 25 genes, 10 genes are described previously in the context of T cell biology and are described in the literature. The following genes are selected due to their association with immune cell biology and signalling. IFIT3 (Interferon induced protein with tetratricopeptide repeats 3) is the highest-ranked protein known as an inhibitor of various cellular processes, including proliferation and signalling; it is also an inhibitor of viral processes promoting antiviral gene transcription (Choi et al., 2018). RSAD2 (Radical S-adenosyl methionine domain containing 2) is also an antiviral protein for mouse dendritic cell maturation (Jang et al., 2018). CD300A is an inhibitory receptor to NK cells and toll-like receptor signalling (Zenarruzabeitia et al., 2015). IFIT1 (Interferon induced protein with tetratricopeptide repeats 1), IFIT2 (Interferon induced protein with tetratricopeptide repeats 2) and MX1 (MX dynamin like GTPase 1) have antiviral functions (Bizzotto et al., 2020; Choi et al., 2018). TGFBR3 is a TGF-B receptor that might be involved in inhibiting TGFB signalling. Decreased TGFBR3 has been

reported to be associated with some cancers (Nishida et al., 2018). SLFN5 (Schlafen family member 5) promotes malignant glioblastoma (Arslan et al., 2017). RXRA (Retinoid X receptor alpha) is a retinoic acid receptor and a transcription factor that regulates gene expression by binding to target response elements, and it has been reported to be expressed in terminally differentiated cytotoxic effector T cells (He et al., 2016). PARP9 (Poly (ADP-ribose) polymerase family member 9) has a role in interferon-mediated immune regulation of macrophages (Iwata et al., 2016).

(A)




(B)

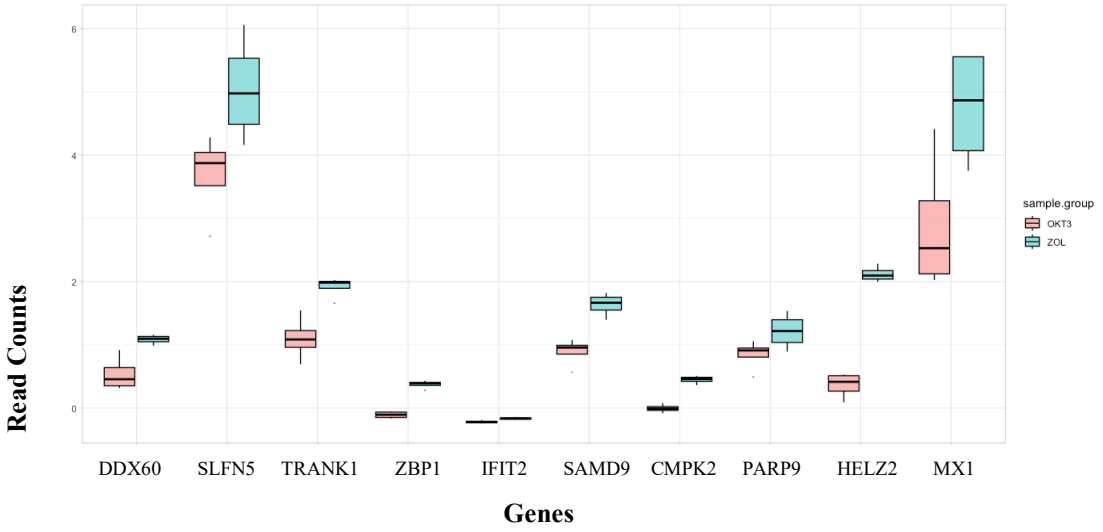



*Figure 5.17: The top 25 significantly downregulated genes in the double stimulated sample are shown in (A) and (B).*

*Functional Analysis*

### Over-representation analysis (ORA): Biological processes

The ORA was also performed in the second experiment to compare and validate the first experiment's results (Figure 5.18). As this data set was acquired with more cells and reads, more significant numbers of gene sets are present in the data.

The top 10 most significant biological processes are shown in the data (Figure 5.18A). While the first experiment resulted in Chemokine and cytokine-based cell regulation biological processes, the second experiment had processes associated with regulating immune cell activation, proliferation, and cytokine production. The only common biological process between both transcriptomics analyses is positive regulation of cytokine production, which is expected as it is essential for regulating immune cells.

Most sets seem to be of a similar size, but the positive regulation of lymphocyte and mononuclear cell proliferation, which contains the smallest number of genes (i.e., 40 genes each)—the most significant sets of mononuclear cell differentiation, leukocyte cell-cell adhesion and cytoplasmic translation. T cell differentiation is a fundamental process in understanding T cell exhaustion and supports that the model developed induced changes by differentiation in the double-stimulated samples.

The second set group is the negative regulation of immune systems, response to virus and regulation of T cell activation. The second group of sets are expected to be present and suggests that chronic stimulation was used, which resulted in a response similar to virus activation and leads to cell activation and negative regulation of some immunity. The GO plot (Figure 5.18A) highlights the relationships and negative and positive regulation of some gene sets to other sets. The cnet network shows specific genes that are associated with each set. In cnet (Figure 5.19), mononuclear cell differentiation highlights essential genes described in T cell exhaustion, including JUNB (proto-oncogene), BATF and TCF7. Leukocyte cell-cell adhesion shows NR4A3. Response to virus set shows IFIT1 and IFIT2, which were mentioned above, have antiviral activity. There is also a set of genes that do not have a known association with biological processes genes (shown as nodes with no title).

The functional analysis shows the general biological processes associated with the double-stimulated samples. The molecular functions set will also be investigated with gene set enrichment analysis of published databases.

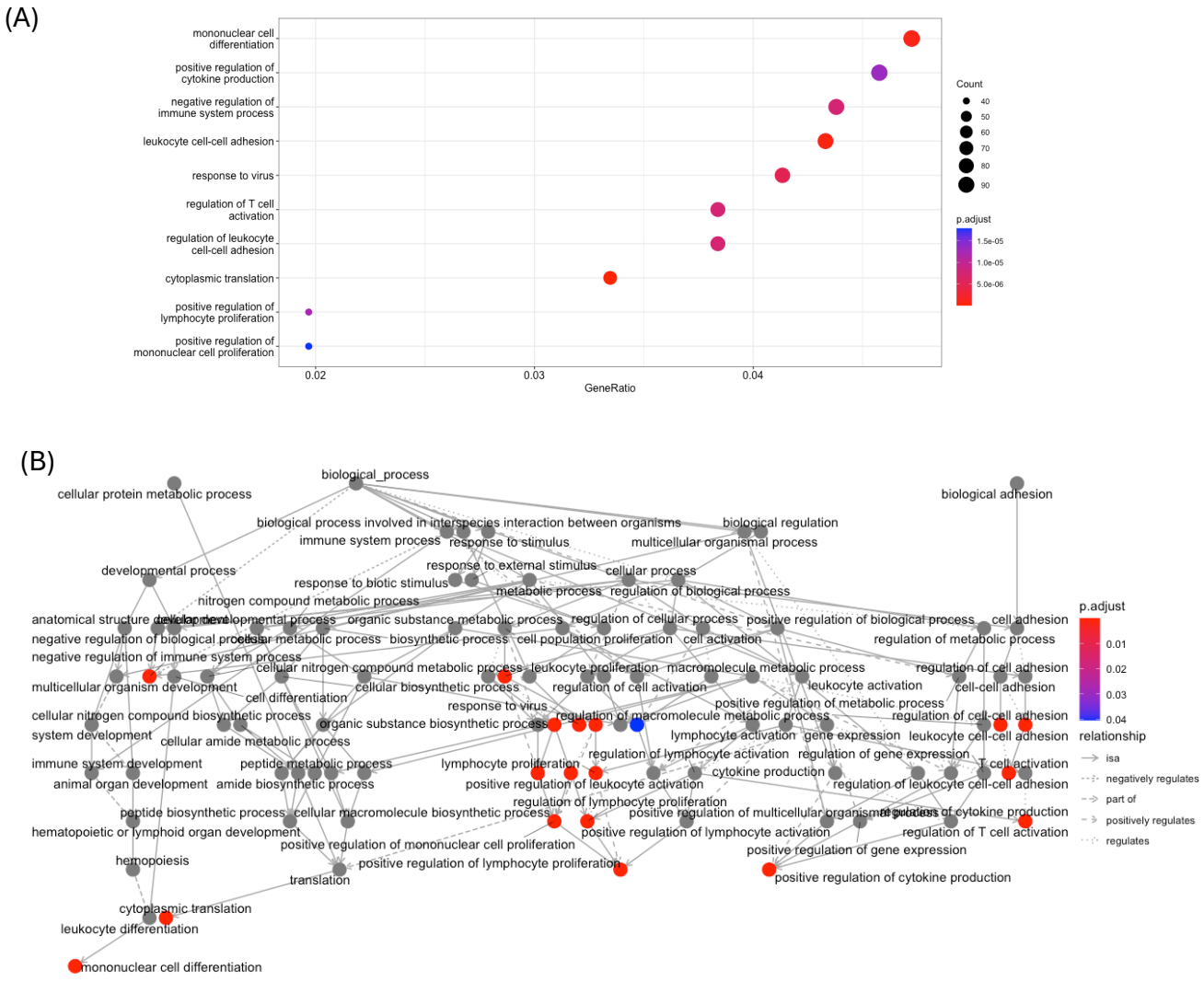



*Figure 5.18: Over-representation analysis in biological process is shown with (A) a dot plot where x-axis is the size of overlap of gene of interest/size of overlap of gene sets collection and dots size corresponds to genes counts in a specific set and (B) a gene ontology network of different biological processes sets and their interactions. All data shown are of significant and with p-adjusted value of <= 0.05.*

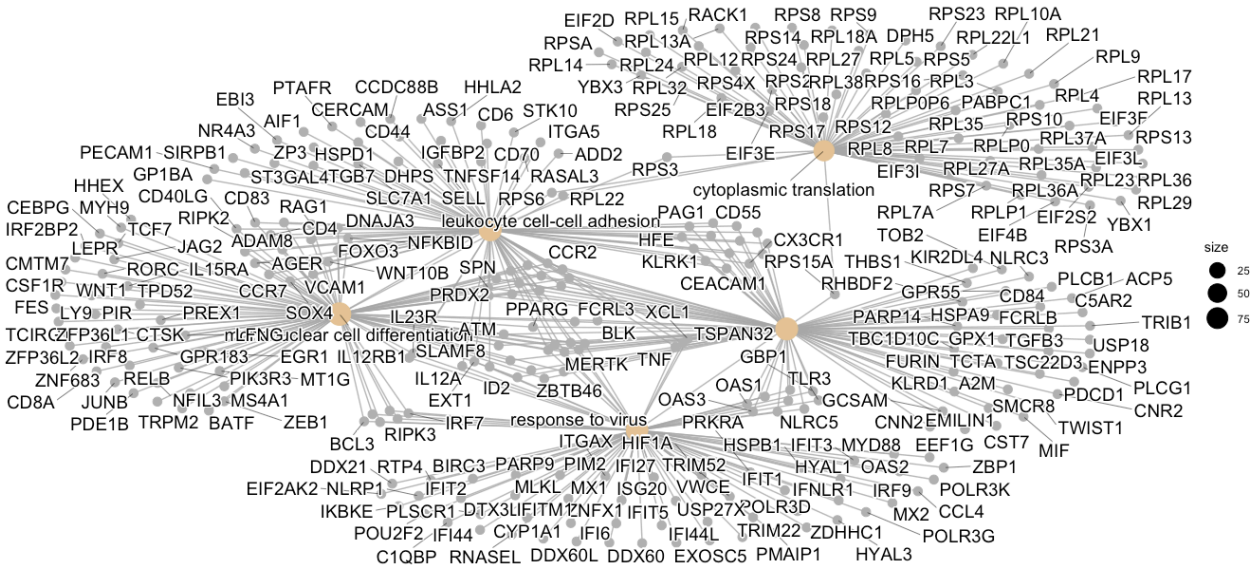



*Figure 5.19: A concept network plot showing all genes from the significant genes list that are contributing to over-representation of each of the biological processes sets. Yellow nodes are used for genesets and greay nodes are for*

## Over-representation analysis (ORA): Molecular functions

The ORA for molecular functions was also performed in this experiment to compare and validate the first experiment's results (Figure 5.20). All data shown are for the most significant processes. As this data set was acquired with more cells and reads, it is expected to notice an increase in the size of gene sets due to the more significant number of genes in the data. All sets are similar to what was shown in the first experiment and include cytokines and chemokines-based processes. Still, the first RNA dataset did not present two groups (relating to a structural component of ribosomes and rRNA binding). The structural constituent of the ribosome was shown as the set containing the highest counts of genes in this analysis. The succeeding groups with the highest counts of genes are cytokine receptor activity, immune receptor activity and cytokine binding. The activation is consistent with the claim that the chronic model activated the γδ T cells and further proved with cytokine receptor activation. The rest of the sets are for chemokine receptor activity and chemokine binding. The GO plot (Figure 5.20B) shows the relationships and does not indicate the presence of a complex network of regulation like what was shown with the

biological processes. However, the relationships are shown as direct lines connecting individual sets that are very similar and do not have any regulatory functions on each other. A Cnet network (Figure 5.21) with listed genes corresponding to the gene sets is also shown.

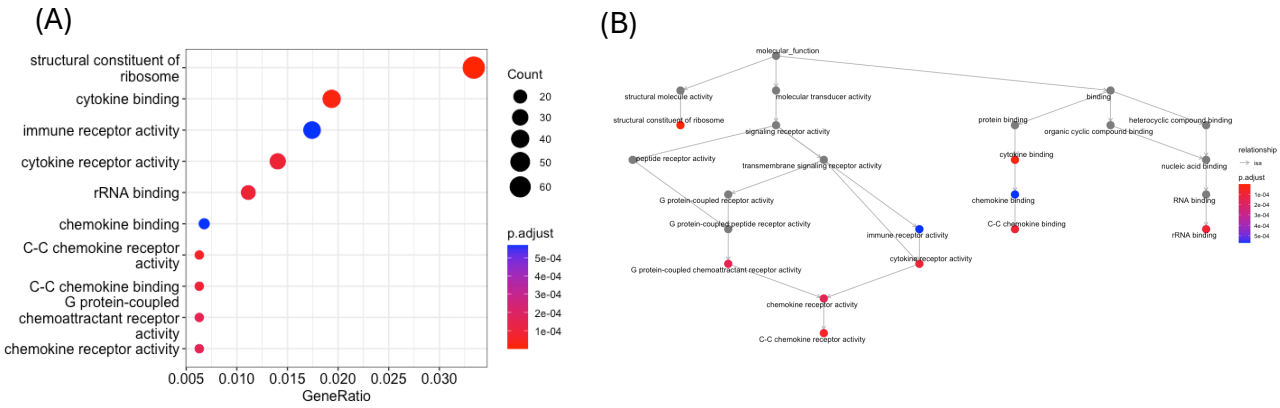



*Figure 5.20: Over-representation analysis in biological process is shown with (A) a dot plot where x-axis is the size of overlap of gene of interest/size of overlap of genesets collection and dots size corresponds to genes counts in a specific set and (B) a gene ontology network of different biological processes sets and their interactions. All data shown are of significant and with p-adjusted value of <= 0.05.*

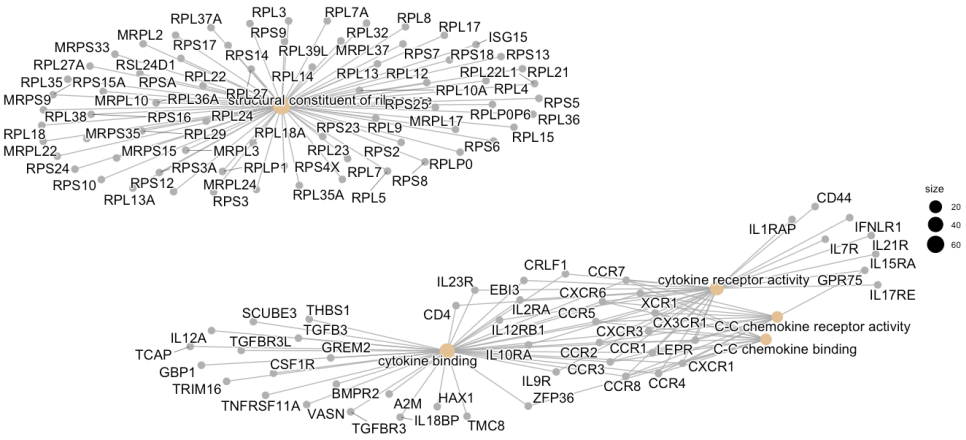



*Figure 5.21: A concept network plot showing all genes from the significant genes list that are contributing to over-representation of each of the biological processes sets. Yellow nodes are used for genesets and greay nodes are for genes.*

*Gene set enrichment analysis: Enrichment maps*

### Hallmark gene sets

A GSEA analysis was then done to show the similarity between published and recent gene sets uploaded to the database.

Gene set enrichment analysis was performed on the complete gene list, and sets with a significant p-adjusted value were shown in an enrichment plot. It is interesting to discover if new gene sets are enriched and different to the first transcriptomic analysis due to the increased number of cells acquired and the use of other donors. The gene set chosen was the hallmarks gene set from the molecular signatures database (msigDB). All the gene sets have a significant enrichment from the gene list of the RNA seq Figure 5.21. Table 5.3 shows the gene sets' corresponding size (number of gene counts), normalized enrichment score (NES) and p.adjusted values.

The hallmark gene sets analysis showed some similarities to the first dataset. The following gene sets were present in both transcriptomic analyses performed and are significant due to their recurrency and association with the double-stimulated Vδ2 T cells condition (OKT3). These sets are the p53 pathway, IL-2 STAT5 signalling, Xenobiotic metabolism, Myc targets, and Glycolysis. The Interferon alpha response and Interferon gamma response gene sets had a negative NES (normalized enrichment score) value in both experiments, which means they are enriched at the bottom of the list for both pathways. However, there were some other gene sets that are specific to the second transcriptomics analysis and include Fatty acid metabolism, Adipogenesis, Oxidative phosphorylation, Spermatogenesis, DNA repair, E2F targets, Estrogen response early, Estrogen response late, Cholesterol homeostasis, MTORC1 signalling, Unfolded protein response and Heme metabolism.

From the second analysis, the unfolded protein response is a cell stress response from the endoplasmic reticulum, and genes found are upregulated during this response. The E2F targets

gene set suggests the presence of genes encoding targets to the E2F transcription factor family involved in cell cycle regulation. Other pathways, such as DNA repair with Heme metabolism and spermatogenesis, are also shown and contain some of the genes expressed in γδ T cells in response to double stimulation. The gene sets generally show multiple pathways involved with cell growth and development, and some sets are specific for stressful conditions or responses to a stimulus. The functional analysis represents potential mechanisms for activating or regulating particular cellular processes in γδ T cells.

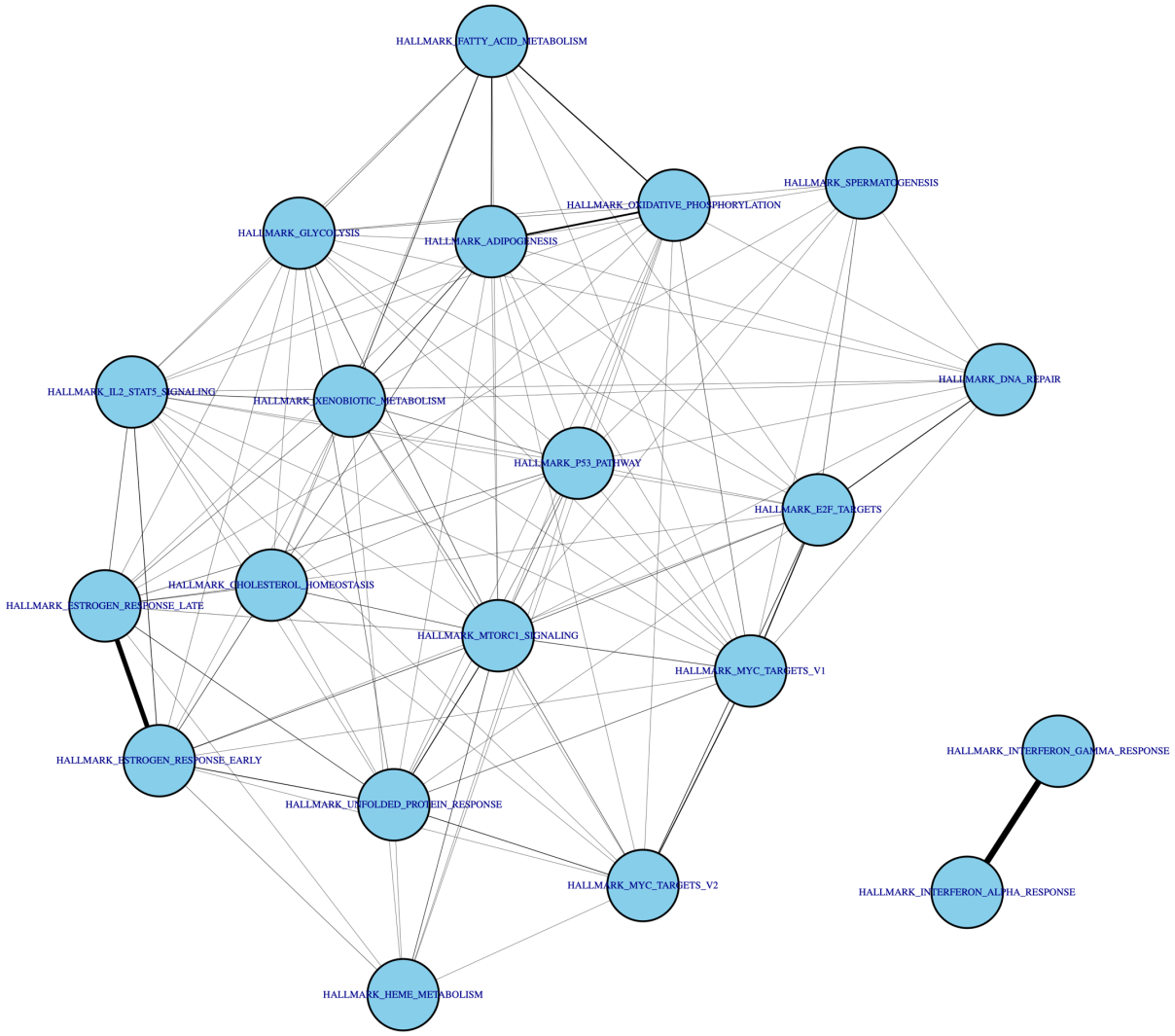



*Figure 5.22: Gene set enrichment analysis of hallmarks genes in the global transcriptome showing significantly enriched sets with p-adjusted <= 0.05.*

*Table 5.3: Enriched hallmark gene sets with set size, Normalized enrichment score (NES) and p.adjust values.*

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **Positively Enriched (with positive NES value)** | | | |
| **HALLMARK_MYC_TARGETS_V1** | 196 | 2.76755389 | 4.30838164381315e-21 |
| **HALLMARK_OXIDATIVE_PHOSPHORYLATION** | 197 | 2.77508582 | 1.8863134757542e-20 |
| **HALLMARK_MTORC1_SIGNALING** | 191 | 2.21984526 | 7.89218462616365e-10 |
| **HALLMARK_UNFOLDED_PROTEIN_RESPONSE** | 108 | 2.31227026 | 1.9792676696499e-08 |
| **HALLMARK_E2F_TARGETS** | 198 | 2.05030262 | 2.57212690266538e-08 |
| **HALLMARK_ADIPOGENESIS** | 164 | 2.10522393 | 7.11793783977157e-08 |
| **HALLMARK_FATTY_ACID_METABOLISM** | 126 | 2.17771843 | 1.83719484779738e-07 |
| **HALLMARK_IL2_STAT5_SIGNALING** | 166 | 1.99672441 | 8.79972062494e-07 |
| **HALLMARK_MYC_TARGETS_V2** | 58 | 2.26944113 | 7.22886249134843e-06 |
| **HALLMARK_CHOLESTEROL_HOMEOSTASIS** | 62 | 2.0391254 | 0.0001068 |
| **HALLMARK_XENOBIOTIC_METABOLISM** | 121 | 1.75405435 | 0.00059564 |
| **HALLMARK_SPERMATOGENESIS** | 74 | 1.88178042 | 0.00061882 |
| **HALLMARK_ESTROGEN_RESPONSE_LATE** | 122 | 1.6646321 | 0.00200163 |
| **HALLMARK_GLYCOLYSIS** | 156 | 1.56780993 | 0.00579 |
| **HALLMARK_DNA_REPAIR** | 141 | 1.52112945 | 0.00770772 |
| **HALLMARK_ESTROGEN_RESPONSE_EARLY** | 128 | 1.45660725 | 0.02579626 |
| **HALLMARK_HEME_METABOLISM** | 155 | 1.40762593 | 0.03234716 |
| **HALLMARK_P53_PATHWAY** | 165 | 1.37404304 | 0.0373828 |
| **Negatively Enriched (with negative NES value)** | | | |
| **HALLMARK_INTERFERON_ALPHA_RESPONSE** | 91 | -2.3172067 | 1.29901080271399e-09 |
| **HALLMARK_INTERFERON_GAMMA_RESPONSE** | 172 | -1.8656701 | 7.40555981084878e-06 |

### Immune-signatures gene sets (C7)

Following the hallmarks set enrichment, it was necessary to check the immune signatures data set. It is more tailored for immune cells and will provide a better overview of the genes expressed in the double-stimulated samples. The complete list of significantly enriched immune-signatures data sets was 1409, with 1018 sets that are less than or equal to 0.01 p.adjusted value.

The immune signatures gene sets (Figure 5.23) show an interesting observation as the top two most significantly enriched sets are of the same study shown with the first experiment. Still, in contrast, it is enriched for upregulated genes between low-dose stimulation with an immune modulatory molecule (GM-CSF) and an agonist of its receptor (Curdlan) and an unstimulated control with high-dose curdlan-only stimulation of DCs (dendritic cell) versus an unstimulated control (Min et al., 2012). T cell-specific gene sets include the downregulated genes when comparing naïve CD4 T cells to 48 or 12 hours stimulated CD4 TH1 cells (Abbas et al., 2005), as well as a comparison between resting and TCR-activated CD4 T cells in a study looking at bystander T cells activation profiles (Bangs et al., 2009) and studies of memory CD4 T cells and TH1 or TH2 cells (Jeffrey et al., 2006).

Genes in double-stimulated γδ T cells profiles showed similarity to genes that downregulated from comparing untreated to IL-6 treated STAT3 knock-out CD4 T cells. The study aimed to identify the role of STAT3 in chronic intestinal inflammation and prove its role in cell function in inflammation (Durant et al., 2010). Another study compared low BCL-6 follicular helper T cells to conventional CD4 T cells. It showed similarity with the downregulated genes from the set to the transcriptome of double-stimulated γδ T cells.
One more gene set was of wild type and TCF-1 deficient CD4 and CD8 double-negative thymocytes (stage 3) where downregulated genes were enriched with signatures of γδ T cells (double-stimulated).

The other group of gene sets is for genes enriched with upregulation signatures between TCR and bystander-activated CD4 T cells (Bangs et al., 2009), Treg versus conventional CD4 T cells from the lymph node of aged mice (Cipolletta et al., 2012), double negative thymocytes versus

double negative thymocytes in a study to identify genes crucial to thymocyte differentiation and lineage determination (Egawa & Littman, 2011). The last gene set is of genes upregulated following a 24-hour TNF stimulation of conventional T cells versus regulatory T cells.

The presence of many different cell types and signatures might suggest the heterogenous culture, which leads to showing signatures of mixed populations of immune cells. Precisely for this study, this analysis aimed to check if there were any enriched signatures of exhaustion-related pathways. For this reason, the gene sets for exhausted genes were picked solely for comparison (Figure 5.24). However, as there are fewer datasets for exhaustion profiles of γδ T cells, it is not optimal to make firm conclusions from comparisons with alpha-beta (αβ) T cells, but it could suggest similar pathways.

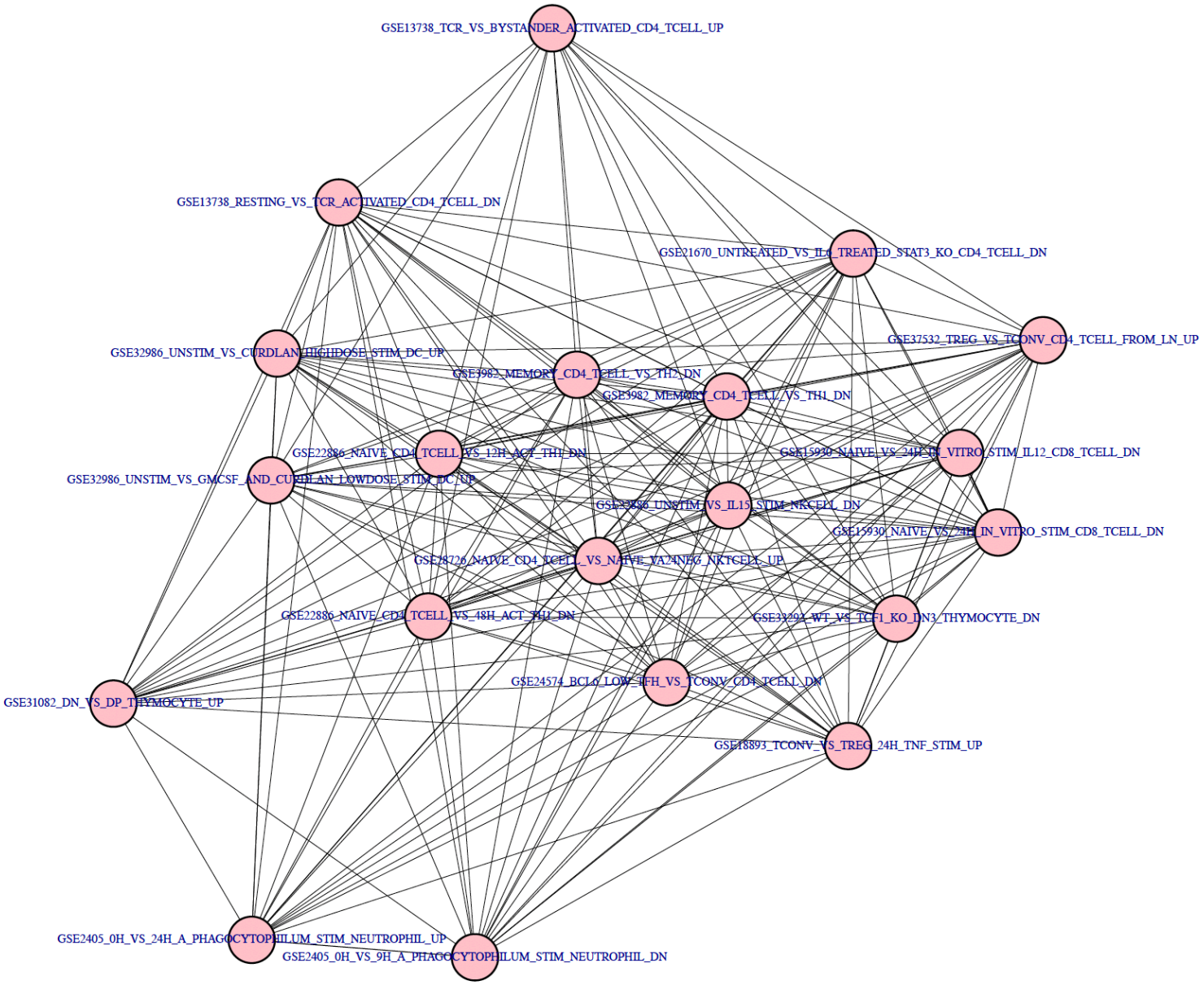



*Figure 5.23: Gene set enrichment analysis of immune signatures genes in the global transcriptome showing 20 of significantly enriched sets with p-adjusted <= 0.05.*

*Table 5.4: Enriched Immune-signature (c7) gene sets with set size, normalized enrichment score (NES) and p.adjust values.*

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **Positively Enriched (with positive NES value)** | | | |
| **GSE32986_UNSTIM_VS_GMCSF_AND_CURDLAN_LOWDOSE_STIM_DC_UP** | 182 | 2.90475109 | 3.55856603097934e-20 |
| **GSE32986_UNSTIM_VS_CURDLAN_HIGHDOSE_STIM_DC_UP** | 187 | 2.77702688 | 3.12079774462021e-17 |
| **GSE2405_0H_VS_9H_A_PHAGOCYTOPHILUM_STIM_NEUTROPHIL_DN** | 195 | 2.73187532 | 5.10230989772942e-17 |
| **GSE22886_NAIVE_CD4_TCELL_VS_48H_ACT_TH1_DN** | 192 | 2.67975677 | 2.36794039579649e-16 |
| **GSE37532_TREG_VS_TCONV_CD4_TCELL_FROM_LN_UP** | 181 | 2.71386325 | 6.9398304945531e-16 |
| **GSE22886_UNSTIM_VS_IL15_STIM_NKCELL_DN** | 192 | 2.64092837 | 1.29484880413165e-15 |
| **GSE28726_NAIVE_CD4_TCELL_VS_NAIVE_VA24NEG_NKTCELL_UP** | 191 | 2.63584169 | 2.89035401143111e-15 |
| **GSE13738_RESTING_VS_TCR_ACTIVATED_CD4_TCELL_DN** | 191 | 2.62172137 | 5.24319152245596e-15 |
| **GSE3982_MEMORY_CD4_TCELL_VS_TH1_DN** | 184 | 2.60382068 | 1.72914652485025e-14 |
| **GSE22886_NAIVE_CD4_TCELL_VS_12H_ACT_TH1_DN** | 194 | 2.56725754 | 1.72914652485025e-14 |
| **GSE21670_UNTREATED_VS_IL6_TREATED_STAT3_KO_CD4_TCELL_DN** | 183 | 2.54419981 | 5.37332022780666e-14 |
| **GSE3982_MEMORY_CD4_TCELL_VS_TH2_DN** | 187 | 2.57330604 | 1.04704894760042e-13 |
| **GSE24574_BCL6_LOW_TFH_VS_TCONV_CD4_TCELL_DN** | 191 | 2.54499372 | 1.33290768894072e-13 |
| **GSE13738_TCR_VS_BYSTANDER_ACTIVATED_CD4_TCELL_UP** | 187 | 2.56009127 | 1.4949976326335e-13 |
| **GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_IL12_CD8_TCELL_DN** | 195 | 2.54576457 | 1.4949976326335e-13 |
| **GSE15930_NAIVE_VS_24H_IN_VITRO_STIM_CD8_TCELL_DN** | 196 | 2.55197371 | 1.5425143104952e-13 |
| **GSE2405_0H_VS_24H_A_PHAGOCYTOPHILUM_STIM_NEUTROPHIL_UP** | 195 | 2.53646884 | 2.15097208235538e-13 |
| **GSE31082_DN_VS_DP_THYMOCYTE_UP** | 187 | 2.53038363 | 4.48905199121519e-13 |
| **GSE18893_TCONV_VS_TREG_24H_TNF_STIM_UP** | 196 | 2.5260499 | 4.90036052012689e-13 |
| **GSE33292_WT_VS_TCF1_KO_DN3_THYMOCYTE_DN** | 186 | 2.54768842 | 5.09093767513744e-13 |

Table 5.5. shows gene sets specific for various γδ T cells activated with different stimuli. Most signatures shown are for unstimulated Vδ1 and Vδ2 T cells, Short-duration stimulated Vδ2 T cells and LPS (lipopolysaccharide)-stimulated Vδ1. The signatures of double-stimulated Vδ2 T cells show similarity to the described gene sets and suggest the presence of heterogeneously

stimulated cells in the cultures tested. In contrast, some might be stimulated due to OKT3, and some are non-responding to OKT3 stimulation but are still in culture.

*Table 5.5: Enriched Immune-signature (c7) for γδ T cells specific gene sets with set size, normalized enrichment score (NES) and p.adjust values.*

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **From RNAseq 1** | | | |
| **GSE22196_HEALTHY_VS_OBESE_MOUSE_SKIN_GAMMADELTA_TCELL_DN** | 168 | 1.65371274 | 0.00393622 |
| **From RNAseq 2** | | | |
| **GSE3720_UNSTIM_VS_PMA_STIM_VD2_GAMMADELTA_TCELL_UP** | 161 | 2.13886748 | 2.47E-07 |
| **GSE27291_0H_VS_6H_STIM_GAMMADELTA_TCELL_DN** | 115 | 1.94608811 | 8.71E-05 |
| **GSE3720_LPS_VS_PMA_STIM_VD1_GAMMADELTA_TCELL_DN** | 156 | 1.71766151 | 0.00063554 |
| **GSE3720_UNSTIM_VS_LPS_STIM_VD2_GAMMADELTA_TCELL_UP** | 152 | 1.70333641 | 0.00115671 |
| **GSE3720_UNSTIM_VS_PMA_STIM_VD1_GAMMADELTA_TCELL_UP** | 165 | 1.64457441 | 0.00195535 |
| **GSE3720_VD1_VS_VD2_GAMMADELTA_TCELL_WITH_PMA_STIM_UP** | 173 | 1.61030801 | 0.0025946 |
| **GSE27291_0H_VS_6H_STIM_GAMMADELTA_TCELL_UP** | 137 | 1.53033822 | 0.01626029 |
| **GSE22196_HEALTHY_VS_OBESE_MOUSE_SKIN_GAMMADELTA_TCELL_DN** | 165 | 1.39558743 | 0.02951225 |
| **GSE3720_VD1_VS_VD2_GAMMADELTA_TCELL_WITH_LPS_STIM_UP** | 168 | 1.37881623 | 0.04306266 |
| **GSE27291_6H_VS_7D_STIM_GAMMADELTA_TCELL_UP** | 158 | 1.37705768 | 0.0431243 |

Table 5.6.  shows T-cell differentiation and exhaustion-related gene sets from both experiments. The main observations are that there are some exhaustion or development-related gene sets, as expected. Still, from the dataset shown, it can be concluded that there is a very heterogeneous Vδ2 population in the cultures. The first notable point is that there are genes similar to upregulated genes in an exhausted population of CD8 T cells compared to memory CD8 T cells (Wherry et al., 2007). The same study found a set of downregulated genes in effector T cells enriched compared to exhausted T cells, suggesting the presence of both an exhausted population and a population of effector T cells. Another study shows that there is an enrichment of downregulated genes in effector T cells when compared to exhausted CD8 T cells with LCMV clone13 infection on day6; however, the same study also shows enrichment of upregulated genes in effector CD8 T cells when compared to exhausted CD8 T cells with LCMV clone13 infection on day8 (GSE41867) (Doering et al., 2012).

For the second experiment, all gene sets from the first experiment are also shown in the second but one (GSE9650_EXHAUSTED_VS_MEMORY_CD8_TCELL_UP). One interesting gene set shows an enriched group of downregulated genes of Naïve and day eight effector CD8 T cells compared to Day 30 exhausted CD8 T cells with LCMV clone13 infection. The same study also shows downregulation of memory, day 6 and day 15 effector T cells compared to day 30 exhausted T cells. Finally, it is demonstrated that genes down-regulated in virus-specific immunodominant (gp33) exhausted CD8 T cells compared to virus-specific subdominant (gp276) cells are also enriched (Wherry et al., 2007).

*Table 5.6: Enriched Immune-signature (c7) for T cell Exhaustion specific gene sets with set size, normalized enrichment score (NES) and p.adjust values*

| Gene Set | Set Size | NES | p.adjust |
|---|---|---|---|
| **From RNAseq1** | | | |
| **GSE41867_DAY8_EFFECTOR_VS_DAY30_EXHAUSTED_CD8_TCELL_LCMV_CLONE13_UP** | 174 | 1.62313052 | 0.00460198 |
| **GSE9650_EFFECTOR_VS_EXHAUSTED_CD8_TCELL_DN** | 129 | 1.55613181 | 0.01338487 |
| **GSE9650_EXHAUSTED_VS_MEMORY_CD8_TCELL_UP** | 133 | 1.58558957 | 0.02364469 |
| **GSE41867_DAY6_EFFECTOR_VS_DAY30_EXHAUSTED_CD8_TCELL_LCMV_CLONE13_DN** | 182 | 1.4152937 | 0.042389 |
| **From RNAseq2** | | | |
| **GSE9650_GP33_VS_GP276_LCMV_SPECIFIC_EXHAUSTED_CD8_TCELL_DN** | 190 | 1.66033023 | 0.00098647 |
| **GSE41867_DAY6_EFFECTOR_VS_DAY30_EXHAUSTED_CD8_TCELL_LCMV_CLONE13_DN** | 185 | 1.61044218 | 0.00119746 |
| **GSE41867_DAY15_EFFECTOR_VS_DAY30_EXHAUSTED_CD8_TCELL_LCMV_CLONE13_DN** | 188 | 1.63321655 | 0.00155605 |
| **GSE41867_MEMORY_VS_EXHAUSTED_CD8_TCELL_DAY30_LCMV_DN** | 170 | 1.56875988 | 0.00538534 |
| **GSE41867_DAY8_EFFECTOR_VS_DAY30_EXHAUSTED_CD8_TCELL_LCMV_CLONE13_UP** | 188 | 1.53436574 | 0.00619952 |
| **GSE9650_EFFECTOR_VS_EXHAUSTED_CD8_TCELL_DN** | 130 | 1.58452156 | 0.01220124 |
| **GSE9650_NAIVE_VS_EXHAUSTED_CD8_TCELL_UP** | 183 | 1.39958243 | 0.01745252 |
| **GSE41867_NAIVE_VS_DAY30_LCMV_CLONE13_EXHAUSTED_CD8_TCELL_UP** | 135 | -1.4749706 | 0.03567789 |

## Shared genes between the two transcriptomics data sets

Figure 5.24 shows the common significantly upregulated genes consistent between the two transcriptomics experiment repeats. Around 20% of genes upregulated in the first experiment were found in the second. A gene weight or percentile calculation was also measured to determine how they differ between both gene sets and their importance. From the upregulated genes, most genes show similar weight in each set, but few are different and have higher ranks in one specific set compared to the other. Figure 5.25 shows the downregulated genes that are common, and almost 50% of the genes found downregulated in the first experiment were also found to be downregulated in the repeat. However, not all seem to have similar weights in both sets, as several genes are different when we compare both together.

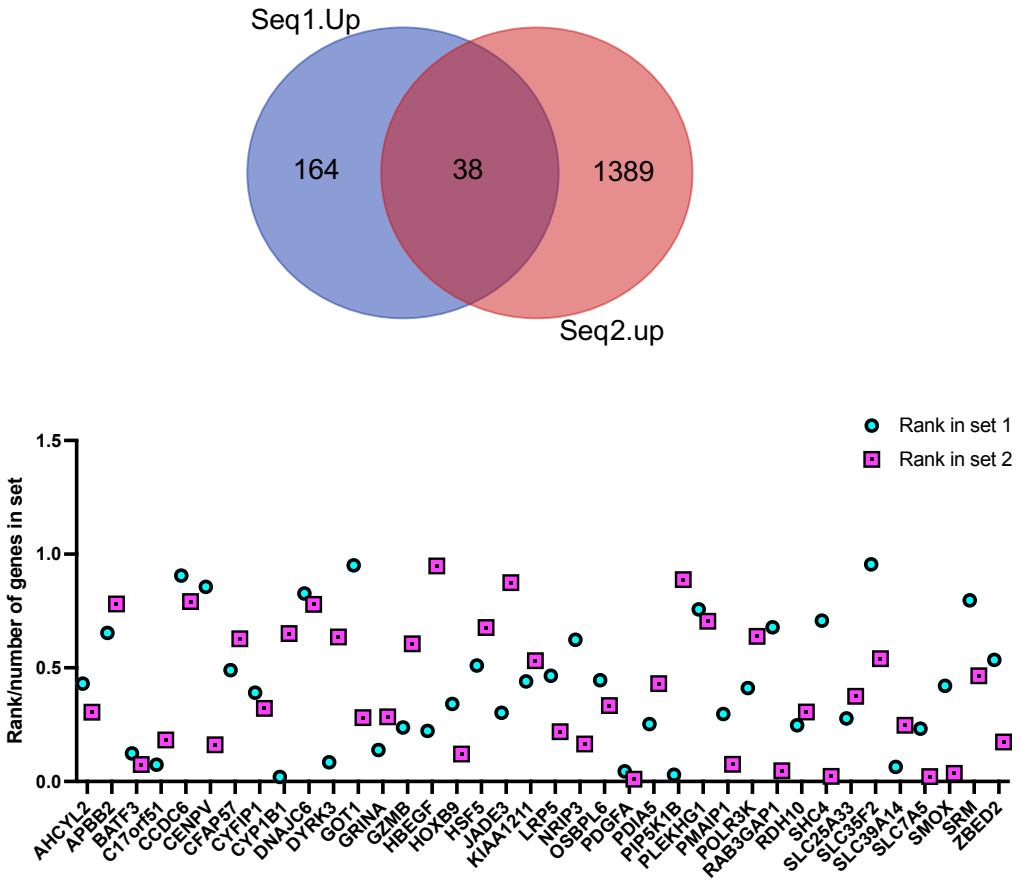



*Figure 5.24: Significantly upregulated genes common between the two repeat experiments are shown with the number of common genes in (A) and gene names with weight of gene in its original list (B). All genes shown are of p-adjusted value <= 0.05. Seq1.Up (or set 1) and Seq2.Up (or set 2) are for experiments 1 and 2, respectively.*

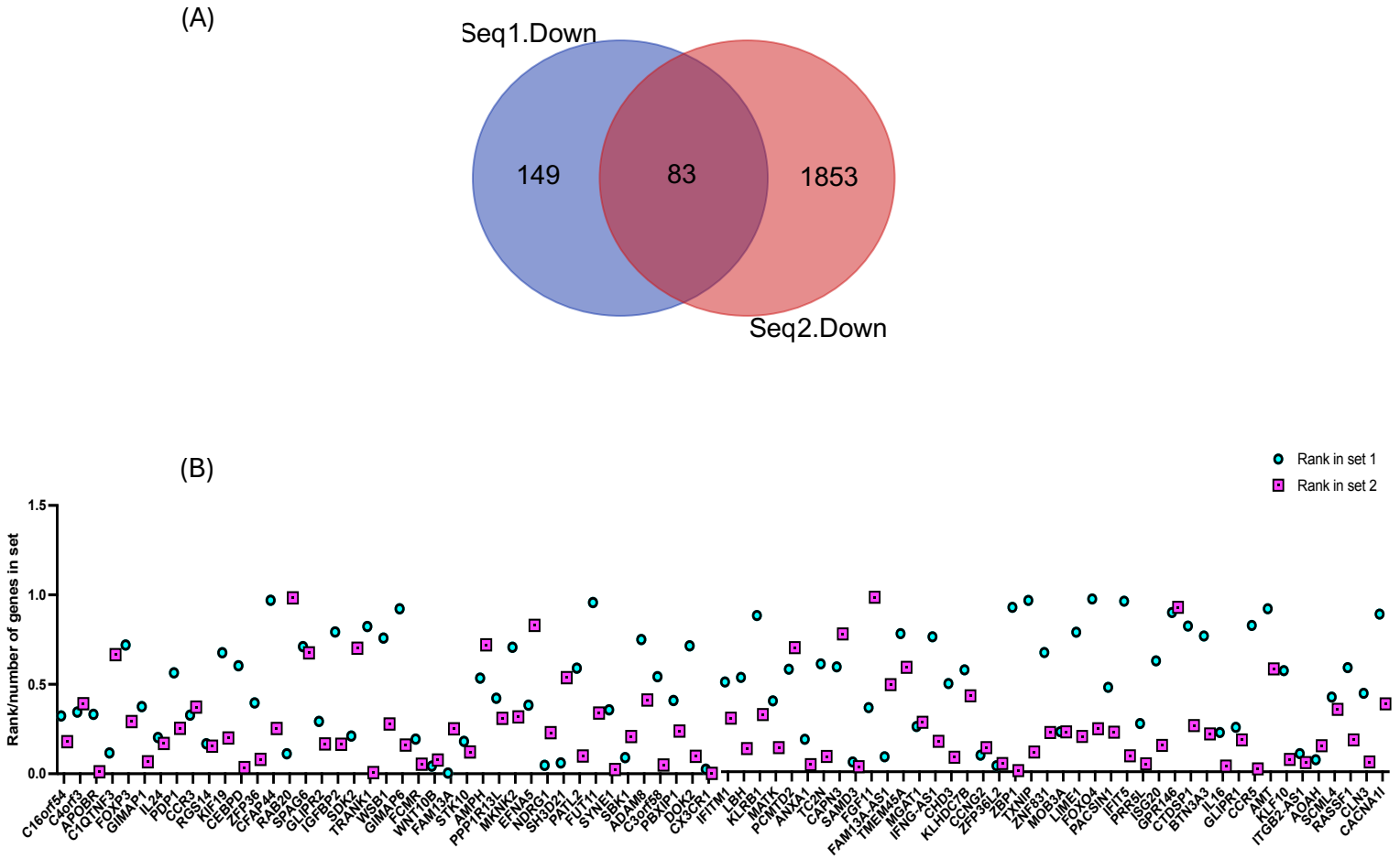



*Figure 5.25: Significantly downregulated genes common between the two repeat experiments are shown with the number of common genes in (A) and gene names with weight of gene in its original list (B). All genes shown are of p-adjusted value <= 0.05. Seq1.Up (or set 1) and Seq2.Up (or set 2) are for experiments 1 and 2,*

## Conclusion

RNA profiling of hypothetically exhausted Vδ2 T cells was performed to validate the phenotypes and characteristics found from the initial experiments with the Zol stimulation model and aim to identify differentially expressed genes in the dual stimulated samples.

The first experiment was done with 3 donors and the profiling showed 202 genes Significantly upregulated and 232 genes downregulated. BATF3 was one gene that is known to be associated with exhaustion and is present in upregulated gene list. IL17F, PDGFA, MAPK11, CD80, CD40LG and CCL22 are known to be related to exhaustion in the literature and were differentially upregulated in the double stimulated Vδ2 T cells.
Downregulated genes that have an impact on T cell development are also present and include MXI1, WNT10B and NDRG1.

A second RNA profiling experiments was done to increase the number of cells collected and include a larger set of donors, further validating the outcomes of the first experiment. For the second experimented, 4 additional donors were profiled for their RNA and analysed. The second experiment show more transcripts as the cell numbers collected was higher than the first time and hence resulted in higher numbers of differentially expressed genes than the initial experiment. Upregulated genes in the dual stimulated Vδ2 T cells included BATF3, LAG3, TOX2, NR4A, TCF4, TCF12 and TCF7L2 are genes that have been described previously in context of exhaustion in conventional αβ T-cells. Differentially downregulated genes were of association to T cells functionality and proteins with inhibitory effects (i.e. IFIT3, RSAD2, CD300A, IFIT1, IFIT2 and MX1 TGFBR3, SLFN5, RXRA and PARP9.

Consideration of both experiments will be made to understand relations between differentially expressed gene and perform analysis on gene networks and their association with Vδ2 T cells exhaustion.

Analysis of both RNA transcriptomics sets shows T cell biology and development gene sets associated with T cell exhaustion gene network expression such as BATF3, LAG3, TOX2, NR4A, TCF4, TCF12 and TCF7L2. The profile of the double-stimulated Vδ2 T cells suggests T

cell differentiation and exhaustion trajectories developments following chronic stimulation on day 6.

Apoptotic pathways are also shown with the genes from the RNA analysis and agree with previous claims of activation-induced cell death following OKT3 stimulation in some populations. On the contrary, cell survival and proliferation signatures are also shown, proving the induction of cell stress and chronic stimulation with the model developed.

Overlap with exhaustion gene sets supports the emergence of exhaustion in the double stimulated Vδ2 T cells. However, it is important to consider that the analysis was of bulk RNA profiliting of a heterogeneous culture. The transcriptome analysis shows the potential presence of cells that are of exhausted signatures in addition to cells that might have experienced activation induced cell death or were of effector and memory phenotypes. Studies with conventional T cells were mostly done with single-cell RNA seq and offered a higher resolution to follow the transcriptome of each cell in the cultures studied. This was not possible in this study due to the high costs of single-cell RNA seq technologies. However, Further links of specific genes from the RNA profiling on γδ T cell functionality will be determined by a pooled CRISPR Knock Out screen.

Our hypothesis is that editing genes that are differentially expressed in the double stimulated samples on day 6 will reverse exhaustion or improve persistence of Vδ2 T cells following chronic stimulation.

# Chapter 6

## CRISPR screens on Vδ2 T cells

### Introduction

#### *Lentiviral vectors*

Viruses are utilised as widespread delivery systems for gene transfer and genetic engineering. Lentiviral vectors are derived from the HIV-1 virus and are single-stranded RNA viruses. They deliver plasmid by integrating DNA in both dividing and non-dividing cells. In library screens such as cDNA or siRNA, using viruses expressing guide RNA there is a requirement for few and preferably only one, integration of DNA per cell. Such viruses are ideal for their use in screens due to their ability to integrate the enriched or dropout sequences into the genome. Consequently, they can be easily detected by sequencing before and after the test conditions. Therefore, a third-generation lentiviral vector system with a CRISPR-Cas 9 system was used for the CRISPR gRNA knockout Library. Third and fourth-generation lentiviruses have been proven safe to use in the clinic, and most known designed CRISPR libraries are constructed in these systems (Ahmadi et al., 2023).

#### *Knockout vs Activation*

CRISPR systems have been recognised for their roles in reprogramming cells by making genetic perturbations in DNA. One of the leading candidates for introducing changes in DNA is the Cas9 system. It recognises a region within the DNA and creates double strand breaks in DNA bases to induce changes such as mutations, insertions or deletions. Following double-strand break (DSB) in the DNA; repairing the DSB with non-homologous end-joining (NHEJ) results in these perturbations that could lead to a loss of function and knockout of the targeted gene, making it an ideal system for knockout screens.

Conversely, a deactivated version of Cas enzymes can also be used to target regions of DNA. Here, instead of causing a total loss of function, it activates or represses the targeted gene by fusing transcriptional activators (i.e. VP64) and repressors (i.e. KRAB) to the deactivated Cas enzyme for gain or partial loss of function studies, respectively. Using the CRISPR system for applications such as overexpression has simplified the need for overexpression library

construction with only the requirement to design a library of guide RNA that is much easier and more affordable.

### *Knockout using CRISPR-Cas9 gRNA Library*

The rationale for a knockout screen is to design a library of selected genes that are candidate mediators of exhaustion or reduced T cell expansion by other non-exhaustion mechanisms. Hence knockout of such genes might lead to preferential enrichment of the guides at the end of expansion in stress conditions. Enriched gRNAs can be indicators of genes that support T cell expansion under prolonged expansion.

### *Considerations for optimising CRISPR screens*

Genome-wide screens can be performed using readily available pre-assembled gRNA libraries targeted to whole genomes of the desired species; however, many cells must be engineered to get adequate coverage of all gRNAs. Targeted CRISPR screens are directed to a limited number of targets selected explicitly for a chosen study. At least four gRNAs for each targeted gene should be included in a CRISPR screen (Bock et al., 2022). In this experiment, five gRNAs were designed per target for 348 target genes. The total gRNA in the gRNA library is 1740.

When performing a pooled CRISPR screen, the cloned gRNA library is usually transfected at a low multiplicity of infection to ensure one gRNA is present per cell. Higher MOIs can be used when cell numbers are restricted, when most of the gRNAs are not expected to affect the cells, or when investigating interactions between genes is required for the study (Bock et al., 2022).

In a screen where positive selection is desired, the coverage needed is 100-200x per target gene selected (Bock et al., 2022). So, the number of cells needed will be roughly calculated by multiplying the desired coverage by the number of target genes (200 (desired coverage) x (348) target genes), which will lead to 69,600 cells being required for the pooled screen. Therefore, for this study, 100,000 cells were chosen for the pooled screen, and a pilot study was performed after optimisations of transducibility of the CRISPR/gRNA plasmid in Vδ2 T cells. Suitable volumes of virus to be used were determined from the optimisations described in this chapter, which included 100ul of concentrated virus, resulting in 30% transduction efficiency with

100,000 cells. Prior to sending the samples for next-generation sequencing, DNA extracted from the cells was amplified around the region of the cloned gRNA library to generate 500bp amplicons. The amount of DNA sent was at a minimum of 500ng per sample and sequencing depth was 10-15 million raw paired-end reads using Illumina MiSeq 2x250 bp.

## CRISPR Knockout Screen on Vδ2 T cells

### *Targets Selection for Knockout Screen*

To construct a list of candidate genes, the top differentially upregulated genes in the double-stimulated samples were selected with a few additional genes associated with the downregulated genes in the double-stimulated sample signatures. Genes associated with downregulated signatures were predicted using Cytoscape software and GENEMANIA application, in which it simulates a gene network from an input gene list (Figure 6.1). In Figure 6.1, downregulated genes are shown in black nodes and represent the downregulated genes while grey nodes represent the genes predicted by Cytoscape to be associated with the downregulated genes, lines indicate interactions between genes and node sizes is proportional to the probable relatability of the gene. Additionally, 11 known published genes that are associated with conventional T cell exhaustion were shown to negatively affect T cells functionality were also included in the CRISPR screen (i.e. TOX1, TOX2, TOX3, NR4A, IRF4, BATF, TIM-3, LAG-3, NFATc1, BLIMP-1 and EOMES) (Bengsch & Wherry, 2015; Kallies et al., 2020; Seo et al., 2019).

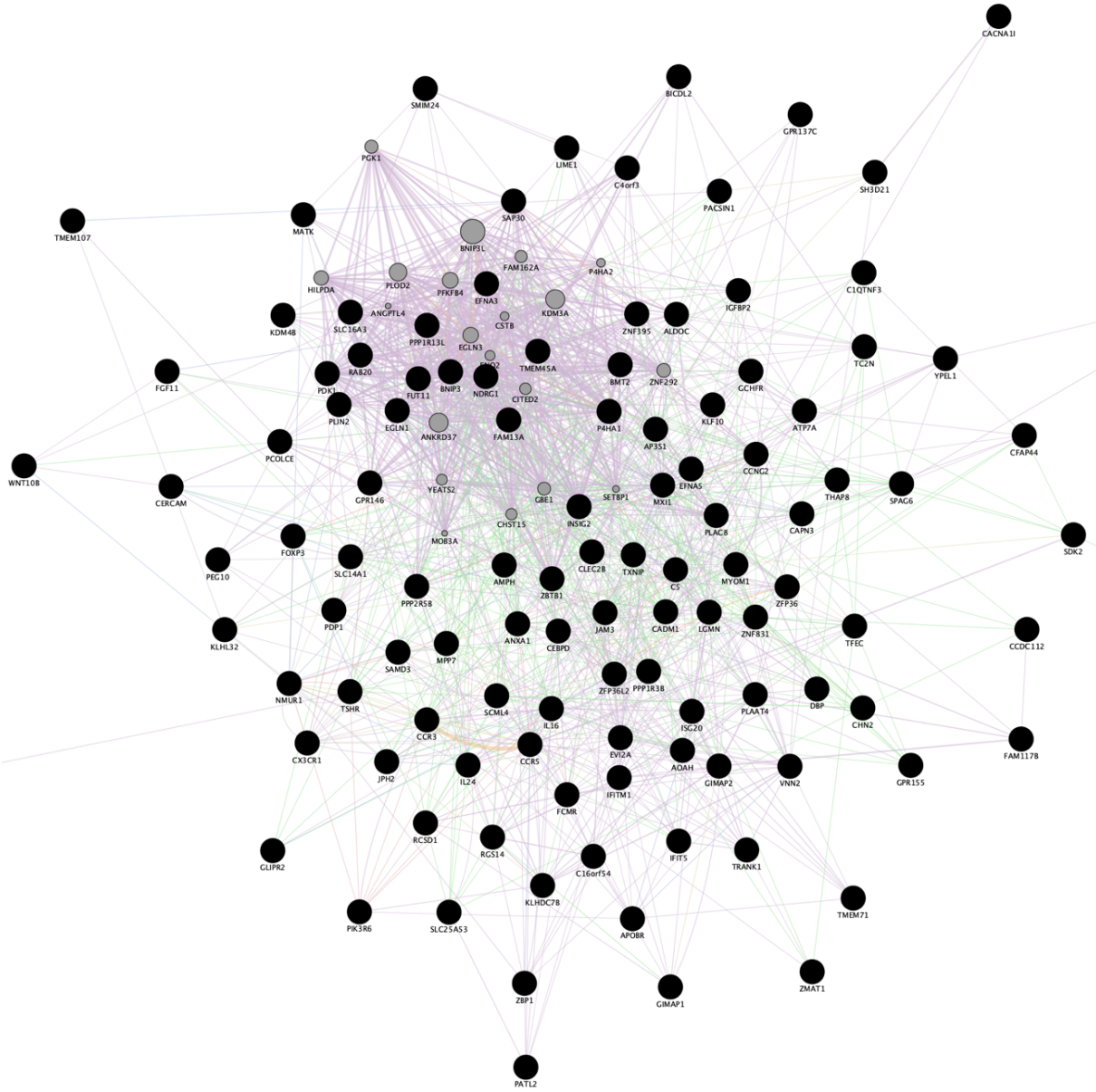



*Figure 6.1: Gene network of downregulated genes in multiple stimulated samples from first RNA profiling (shown as black nodes) with genes of predicted associations to the downregulated genes (shown as grey nodes).*

### *Cas9 System and gRNAs Design*

Several CRISPR systems were considered for the screen. As many publications validate the CRISPR Cas9 system (Stadtmauer et al., 2020; Zhou et al., 2023), it was chosen and ordered as a vector with Cas9 enzyme, a region to insert selected gRNAs through cloning and with a fluorescent protein GFP pLentiCRISPR-EGFP was a gift from Beat Bornhauser (Addgene plasmid # 75159; http://n2t.net/addgene:75159; RRID: Addgene_75159) (McComb et al., 2016). The gRNA insertion site is located between a U6 promoter and a gRNA scaffold region. The U6 promoter is widely used to express CRISPR gRNAs widely in addition to a gRNA scaffold region that is important to bind the Cas9 enzyme. The Cas 9 enzyme is expressed from the same vector, with transcription being regulated by an EF-1 α (Human elongation factor-1 alpha) promoter that is known to be potent and drive high gene expression.

CRISPick software was used to design all gRNAs for chosen targets. Five gRNAs were designed per target and selected by the highest knockout scores and low off-target effects as suggested by the software algorithm.

TWIST bioscience was chosen to produce the library and clone the gRNAs into the vector, as they have expertise in CRISPR library designs. The distribution and coverage of cloned gRNAs by NGS (next-generation sequencing) are shown in Figure 6.2. The cloned library was provided as a DNA aliquot and glycerol stocks.

To expand the library upon its receipt, bacteria from glycerol stocks were grown in LB (Lysogeny broth) cultures overnight with ampicillin to selectively expand the CRISPR library vector. Next day, large-scale tissue culture-grade DNA extraction was performed. The pooled library DNA was then used for virus production.

### *NGS Coverage and Distribution Histogram of KO CRISPR Library*

Distribution histograms show the range and uniformity of sequencing coverage to a data set. The sequencing reads (x-axis) are aligned to a reference (y-axis). Mapped reads are shown in a Poisson-like distribution, expressing the probability of several events existing in a fixed time with a small standard deviation. A narrow-range histogram confirms good sequencing coverage and a broad-range histogram suggests poor sequencing coverage of the requested bases substitutions. The distribution histogram (Figure 6.2) confirms the overall uniformity and reliability of the cloned library as shown by the narrow range sequencing coverage.

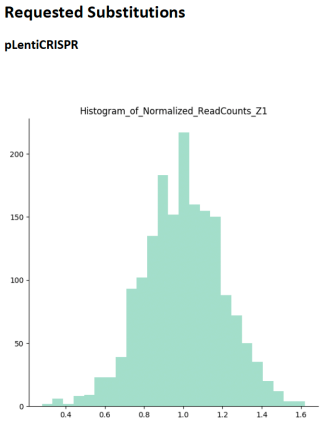



*Figure 6.2: Mapped read depth in CRISPR screen by TWIST Bioscience.*

Currently, no literature exists on using CRISPR screen lentiviruses on Vδ2 T cells. Therefore, a new protocol had first to be developed and optimised to obtain optimal virus titers on Vδ2 T cells. Tested methods were the use of different transfection agents (GeneJuice and PEI), evaluating the effects of spinfection in increasing transduction efficiencies, testing two different virus packaging envelops (VSV-g and RDpro) and finally evaluating the effect of using retronectin (transfection agent) to boost viral titers.

## Use of Different Transfection Agents to Increase Viral Titres

### *GeneJuice vs PEI on Lenti X 293Tcells*

Different transfection agents were tested for their efficiency in generating the vesicular stomatitis virus G (VSV-G) protein pseudotyped lentiviral vector. Polyethylenimine (PEI) is a synthetic cationic polymer commonly used for transfecting cells with DNA and is easily made and widely accessible. However, as PEI must be made within a research lab setting, its preparation method could introduce batch variability and affect transfection efficiency between preparations. On the other hand, GeneJuice is a cationic lipid that forms a liposome with negatively charged DNA and allows transfer to the cells. Lipid transfections are known to be of high efficiency, low variability and applicable to a wide range of cell types. For comparative reasons, PEI was tested against GeneJuice for the CRISPR pool DNA to transfect Lenti X 293T cells. Harvests at 48 and 72 hrs were tested individually to determine variations between the transfectants. As shown in Figure 6.3. Harvests from 48 hours have higher transfection efficiencies with Lenti X 293T cells, with GeneJuice superior to PEI with 70.5% GFP positive cells compared to 33.3% with PEI. The

second harvest at 72 hours showed reduced transfections with similar efficiencies between GeneJuice and PEI with 16.3% GFP cells and 19.5% GFP cells, respectively.

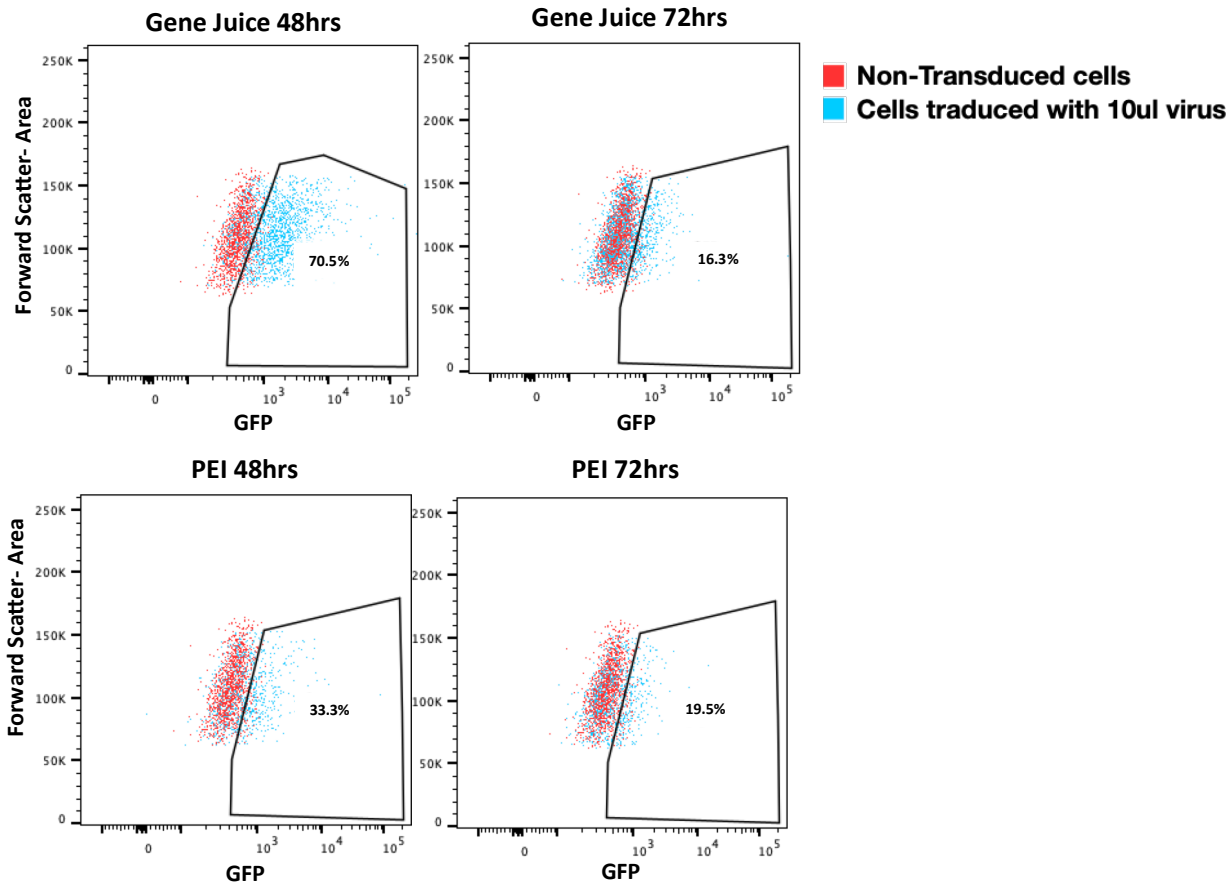



*Figure 6.3: Transduction comparison of GeneJuice(Top) and PEI (Bottom) on Lenti X 293T cell lines after 48 or 72 hrs from transfection.*

## *Effects of Spinfection on Increasing Viral Titres*

The effects of spinfection were studied to show if it has any enhancing effects on virus transductions. Spinfection also known as spinoculation is a method that is reported to increase the transductions of viruses (Rajabzadeh et al., 2021). Figure 6.4 shows that the use of spinfection to slightly increases the transfection efficiency of the concentrated virus on Lenti X 293T cells.

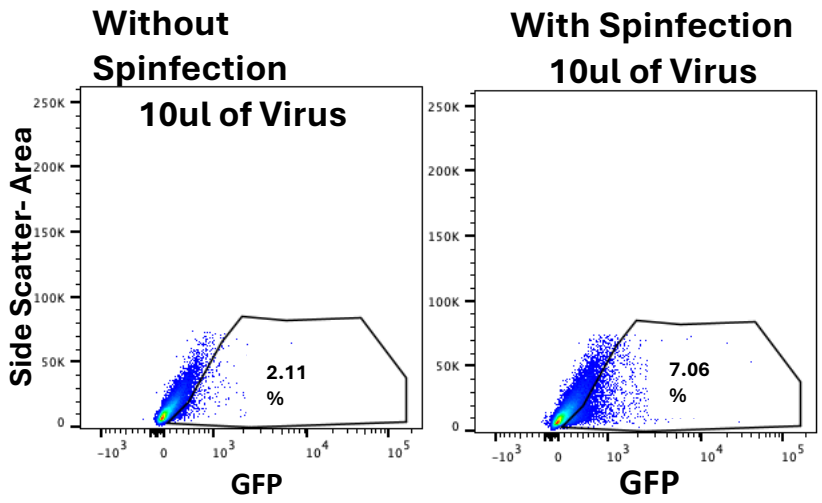


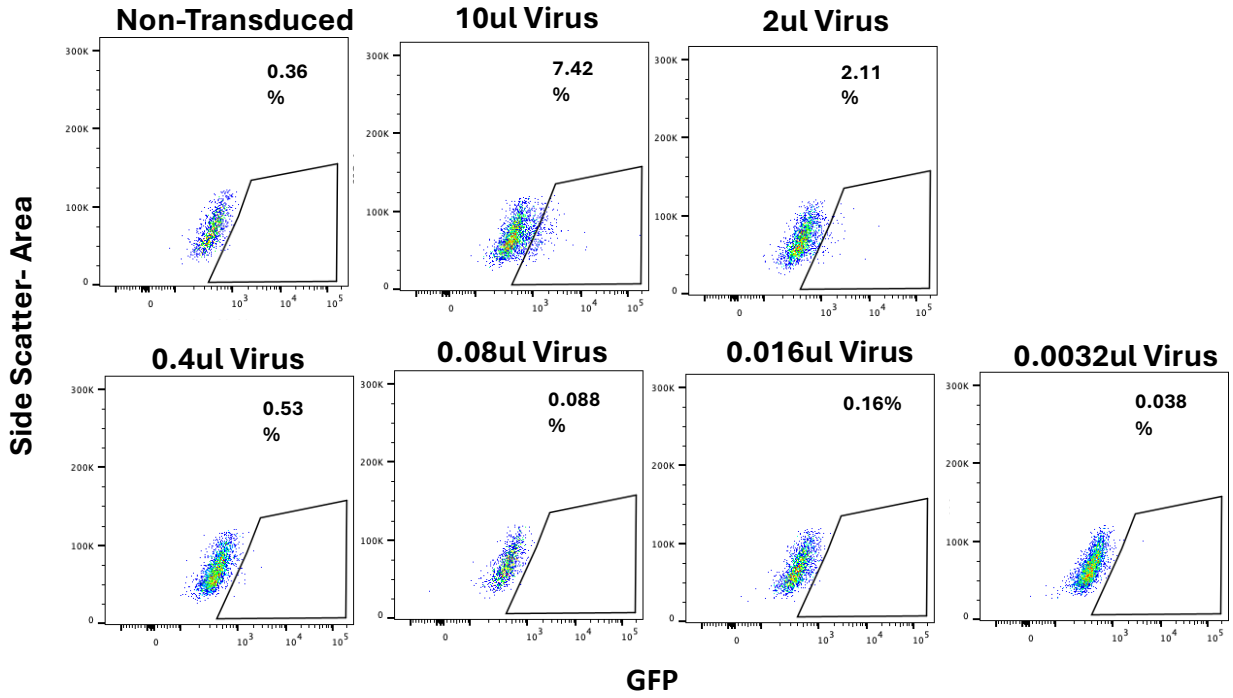

*Figure 6.4: Use of spinfection to increase viral titres on Lenti X 293T cells.* *Fig* *infection.*

### *Impact of Different Viral Envelopes on Transducing Vδ2 T cells*

Virus pseudotyping is important for Cell transfections and engineering. To ensure DNA transfer with high efficiency to any cell type, the appropriate pseudotype envelope is required to ensure viral internalisation into the desired cells. The vesicular stomatitis virus G (VSV-G) protein pseudotyped lentiviral vector and feline endogenous gamma retrovirus RD114 are commonly used for T cell engineering. An improved gamma retrovirus RD114 with a modified cytoplasmic tail has also recently been used for enhanced transductions in T cells and is called RDpro (Tijani et al., 2018).

Comparing two envelopes (VSV-g and RDpro) for titration on Vδ2 T cells resulted in 7% transduced cells with the RDpro enveloped virus as indicated by GFP positivity in Figure 6.5 for the first dilution (D1), which was 10 ul. As expected, the lower dilutions from Figure 6.5 showed much lower transductions on Vδ2 T cells. However, using the same dilutions of the VSV-g virus (Figure 5) showed no transductions with all dilutions tested. This comparison shows that RDpro pseudotyping is better suited to transduce Vδ2 T cells. The VSV-g envelope is commonly used but not ideal for Vδ2 T cells; thus, all the experiments for the CRISPR screen will be performed using the virus with RDpro envelop.

**Titrations on Vδ2 (RDpro envelope)**




**Titrations on Vδ2 (VSV-g envelope)**

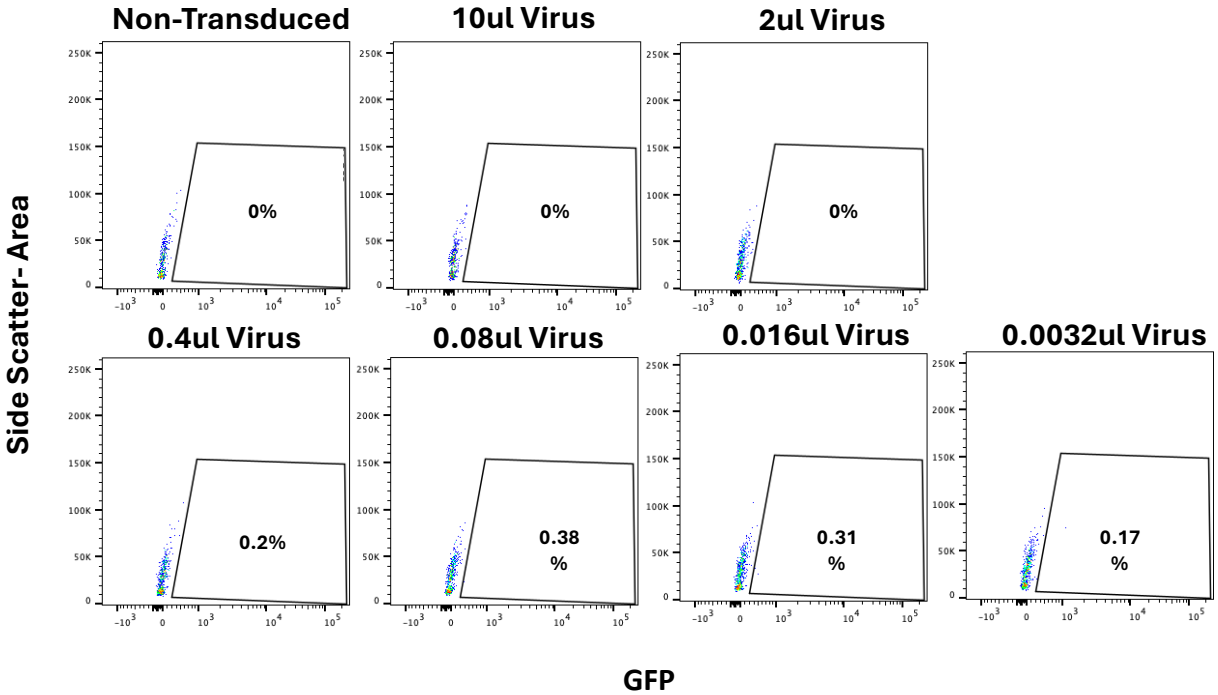



*Figure 6.5: Viral titrations on Vδ2 T cells using virus packaged with RDpro envelope (Top) and VSV-g envelope (Bottom).*

*Retronectin Use to Boost Viral Titers*

The following study tested the RDpro envelope with retronectin transfection reagent and spinfection technique for its use on γδ T cells. Spinfection showed improved transductions in lenti X cell lines (Figure 6.4) while RDpro in Figure 6.5 had better transduction efficiencies when compared to VSV-g on Vδ2 T cells. Retronectin (Takara) is a reagent commonly used with viruses, it consists of fibronectin fragments and is developed to enhance transductions and transfections of hard-to-transfect cells including hematopoietic cells by co-localisation of virus and target cells. Retronectin has 3 domains: a cell binding domain via VLA-5 integrin receptor (fibronectin C-domain), a virus binding domain (heparin H-domain) and a VLA-4 integrin binding site (named as CS-1 sequence). For this experiment, the effects of combining RDpro, retronectin and spinfection was evaluated on Vδ2 T cells. As Vδ2 T cells are hard to transfect, a test with 20ul and 50ul of virus was done with retronectin-coated plates and centrifugation of the virus for 2 hrs at 37C to enhance transfection efficiency and contact of the virus with cells. The required number of GFP cells was achieved using 50ul of the virus, as shown in Figure 6.6, with around 14% of GFP cells (excluding the GFP percentage of non-transduced cells). For a pooled CRISPR library, transfection must be maintained between 10-30%. Following this test, 100ul virus was chosen as the appropriate volume for the CRISPR screen knockout screen to ensure maximum efficiency of around 30% transduced cells.

**Day 4 analysis of RDpro and retronectin with spinfection titrations on** Vδ2

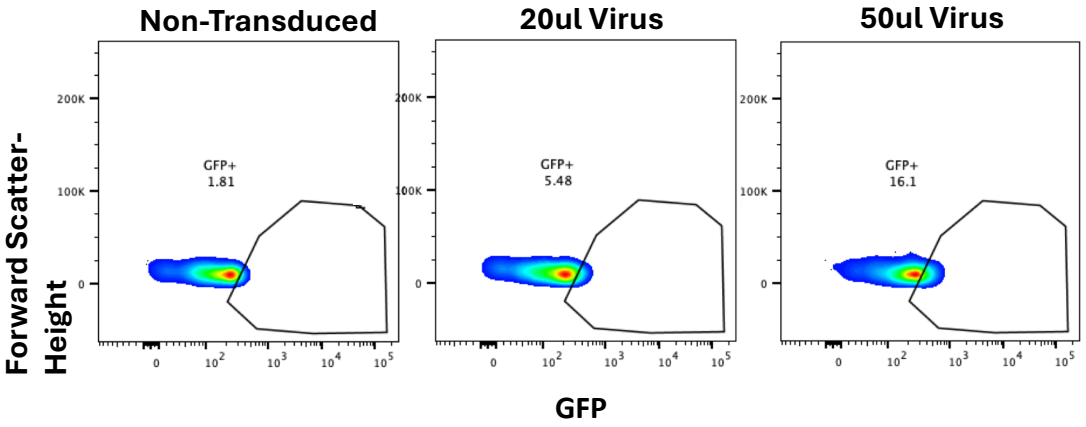



*Figure 6.6: Day 4 Viral titration results on Vδ2 T cells of RDpro enveloped virus with retronectin agent and spinfection.*

In summary, from all the comparisons, GeneJuice was proved to be the best transfection agent to produce the virus using Lenti X 293T cell lines. At the same time, the use of retronectin transfection with Vδ2 also improved the transduction efficiency when followed by a spinfection step. The use of RDpro envelope was clearly shown to have better effects from the VSV-g as shown with lentix 293T cells. So, the overall best method for high titer viral transduction in Vδ2 T cells was using a virus packaged with RDpro in retronectin pre-coated plates and followed by a spinfection, as shown with the test in Figure 6.6, which is the method that was selected for all upcoming experiments.

## Pilot Expansion for Increased GFP Enrichment Following Transduction

Pooled CRISPR screens are widely used to evaluate potential gene targets associated with cell functions, specifically in T-cell Exhaustion. Many reports have investigated exhaustion development in T cells in cancer and chronic infections and prove the impact of such studies in understanding biological processes and interesting targets for therapies (Belk et al., 2022; Legut et al., 2022; Zhou et al., 2023). Following the virus production of the pooled CRISPR knockout screen, transduction of γδ T cells was performed according to the timeline shown in Figure 6.7. Gamma delta (γδ) T cells were activated and expanded as described before in the methods section 1.1 with ZOL and IL-2. Virus transduction from section 4.4 in the methods was followed with the following refinements: On day 9, the cells were stimulated with the first OKT3 (1ug/ml) and again on day 12 then finally on day 15. All conditions were left in culture for four days and collected for staining and preparation for DNA extraction.

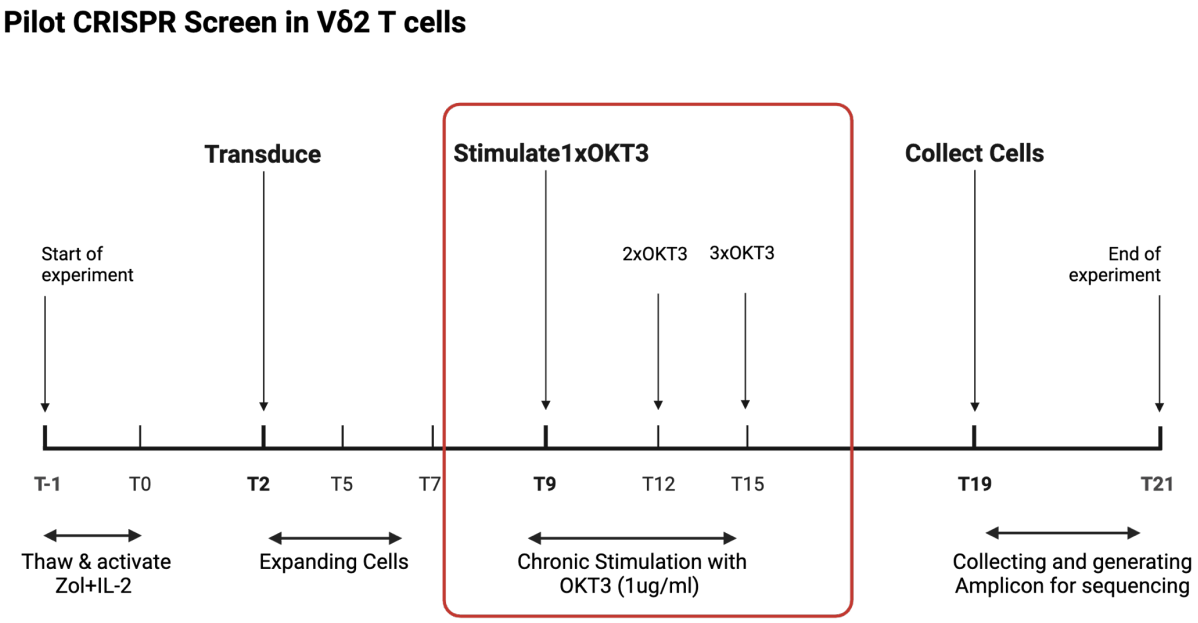



*Figure 6.7: A time-line for pilot CRISPR screen assay.*

On day 9, before starting OKT3 stimulations, the cells were phenotyped for their Vδ2 T cell proportion and GFP (Figure 6.8). All donors had different Vδ2 T cell proportions; donors 1 and 2 were 10-15% GFP positive, while donor 3 was around 50% GFP, possibly due to the higher proportion of the Vδ2 T cell population, consistent with a higher proliferative rate of Vδ2 in that individual leading to higher transducibility. After the additional OKT3 stimulations (adding OKT3 on days 9 and 12), the cells were phenotyped on day 19 for their GFP expression (Figure 6.8). The cells showed 20% GFP expression for donor 2 and 40% GFP for donors 1 and 3. The pilot experiment validates the success of viral transduction using the CRISPR screen knockout library on Vδ2 T cells. This data from the pilot model was used in design of a definitive model accommodating more conditions to be tested and incorporating additional phenotyping of Vδ2 T cells and GFP before and after initiating the differential stimulation.

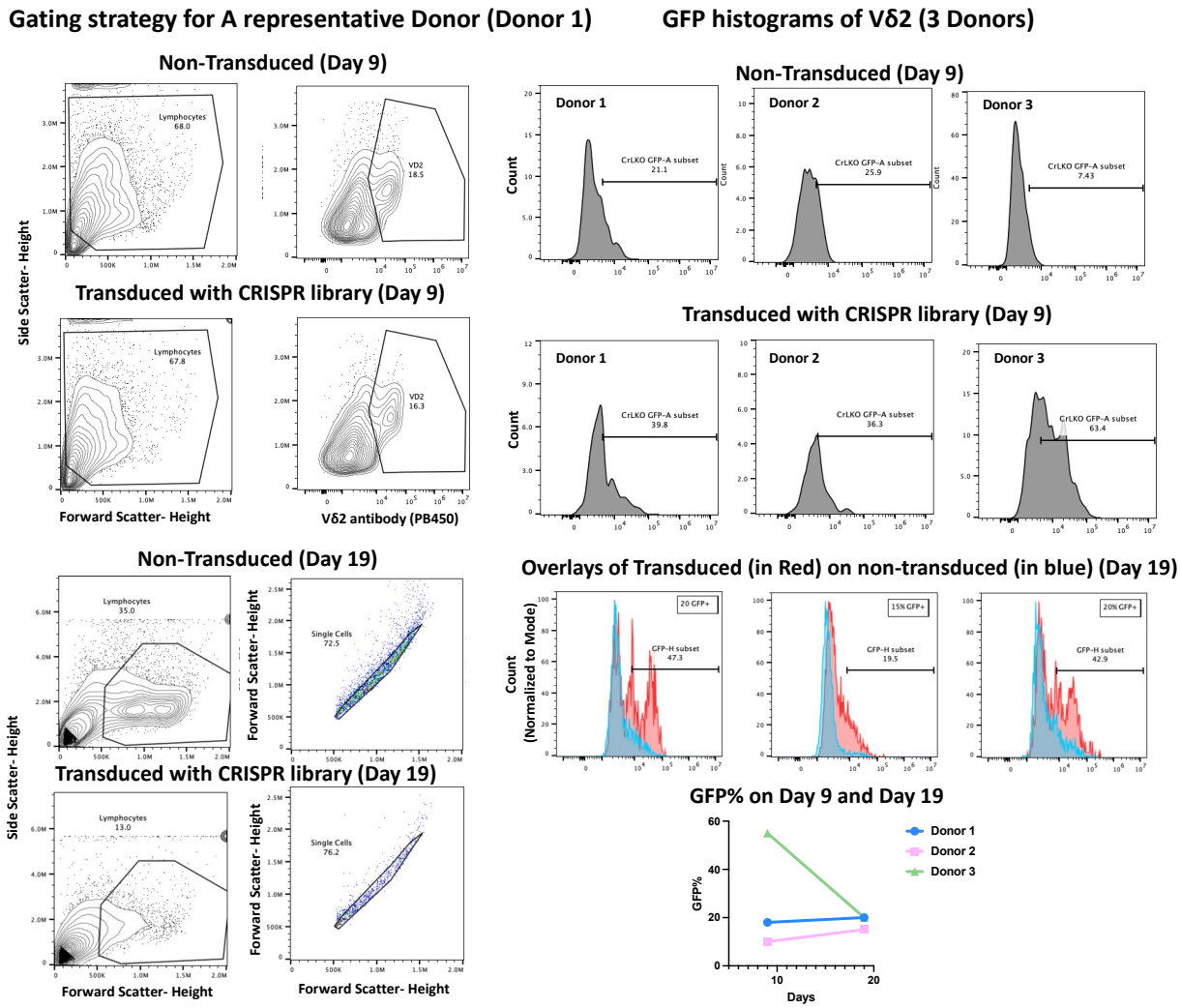



*Figure 6.8: FACS analysis on Day 9 and Day 19 post ZOL expansion of Vδ2 T cells transduced with CRISPR Library. Three donors are shown with representative gating plots and histograms to show GFP positive subsets of Vδ2 T cells on Day 9 and lymphocytes on Day 19 comparing the non-transduced control to the transduced. GFP percentage differences between Day 9 and Day 19 are also shown with a line plot for all 3 donors.*

**Conclusion on pilot Vδ2 cells transduction**

The pilot study showed that the virus used was sufficient for transducing Vδ2 T cells, it showed that transductions of Vδ2 T cells were successful on Day 9 (after ZOL expansion and before any stimulation with OKT3) and remained GFP-positive on Day 19 (following 1 or 2 times OKT3 stimulations) (Figure 6.8). A similar time-line will be implemented on the final CRISPR screen study but additional modifications will be added to ensure enough cells are maintained to study all desired conditions to test for the CRISPR screen.

## Expansion with optimised time-line

Following the pilot data, a modified CRISPR screen assay was designed using a similar timeline for stimulations (shown in Figure 6.9) but with starting OKT 3 stimulations on Day 12 and separating different stimulations every other day. Also, a condition with a 3rd OKT3 stimulation and a condition with adding Daudi target cells at a 5:1 E: T ratio of Daudi to Vδ2 T cells were added. The stimulation was provided on Day 12, day 14 and Day 16 from ZOL expansion for 1x, 2x or 3x OKT3 conditions. The Daudi stimulus was added on day 12 from the day of ZOL expansion. All the conditions were collected on day 19 and prepared for subsequent analysis for GFP expression and pellets for DNA extraction and amplicon generation for sequencing.

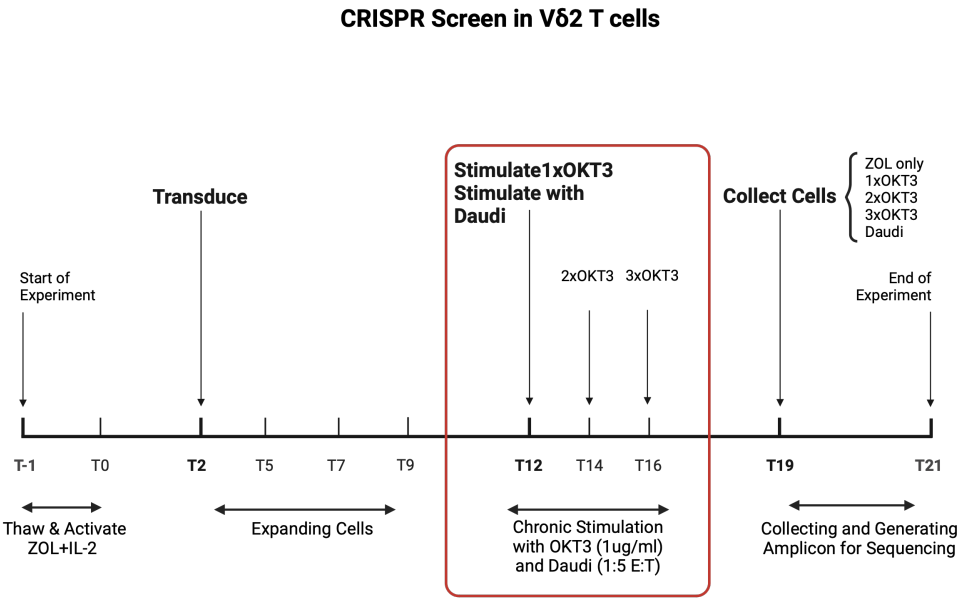



*Figure 6.9: An experimental time-line for CRISPR screen in Vδ2 T cells.*

Two donors had 17 and 13 % GFP-positive cells on day 12 (Figure 6.10), indicating successful transduction of the CRISPR library. The third donor had no GFP cells, indicating failure of transduction that may be due to the inappropriate use of the virus or just the inability of this donor to be transduced. All three donors had more 50% Vδ2 population on day 12. The analysis on day 19 included 5 different conditions per donor: ZOL only, 1x OKT3, 2x OKT3, 3xOKT3 and Daudi-stimulated Vδ2 T cells. All the stimulation follows ZOL on Day -1 (shown in Figure 6.9). Donor 1 (LP-2) had the most GFP enrichments on day 19 (day of collection), where enrichment of GFP cells was noticed from ZOL to 3xOKT3 conditions. The population was increasing slightly with additional stimulation of OKT3 (Figure 6.11).

However, it was noted that there was a background of GFP expression in all conditions, with a large increase in background GFP fluorescence with Daudi stimulation for all donors. considering the GFP fluorescence background, the Daudi conditions still resulted in the lowest GFP-positive populations in all donors. Donor 2 (BC-E) had around 10% GFP-positive cells for the OKT3 Conditions but very few (less than 5%) GFP cells with the ZOL only and Daudi conditions (Figure 6.11). Donor 3 (BC-D) had around 30% GFP for all OKT3 conditions but only 8% for the ZOL-only sample and 5% for the Daudi sample. From the analysis, Donor 2 seems to have the lowest transduction of the CRISPR library, while LP-2 has the most transduction efficiency with increasing enrichment with the different conditions. All Daudi conditions seem poorly transduced as shown in the percentages in Figure 6.11. The differences in levels of transductions could be due to donor variability as different Vδ2 donors can have varied susceptibility to transductions. The samples were all processed for DNA extraction and PCR amplification of the gRNA containing region before being sent for further analysis by next-generation sequencing.

## FACS on Day 12 (T12) before stimulation with OKT3 or Daudi

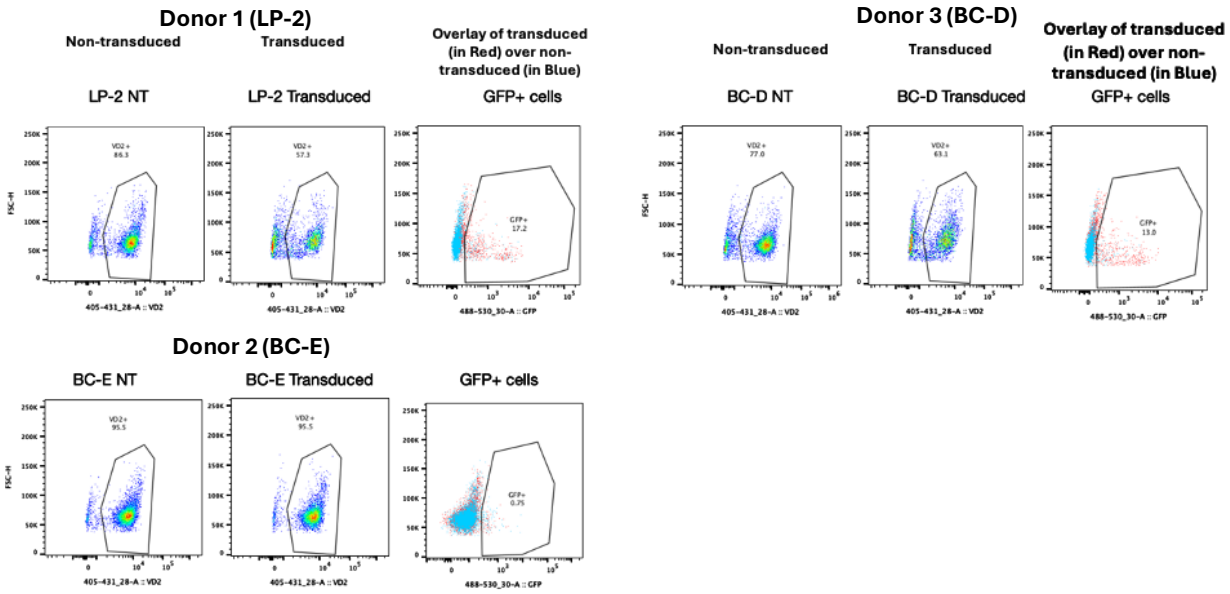
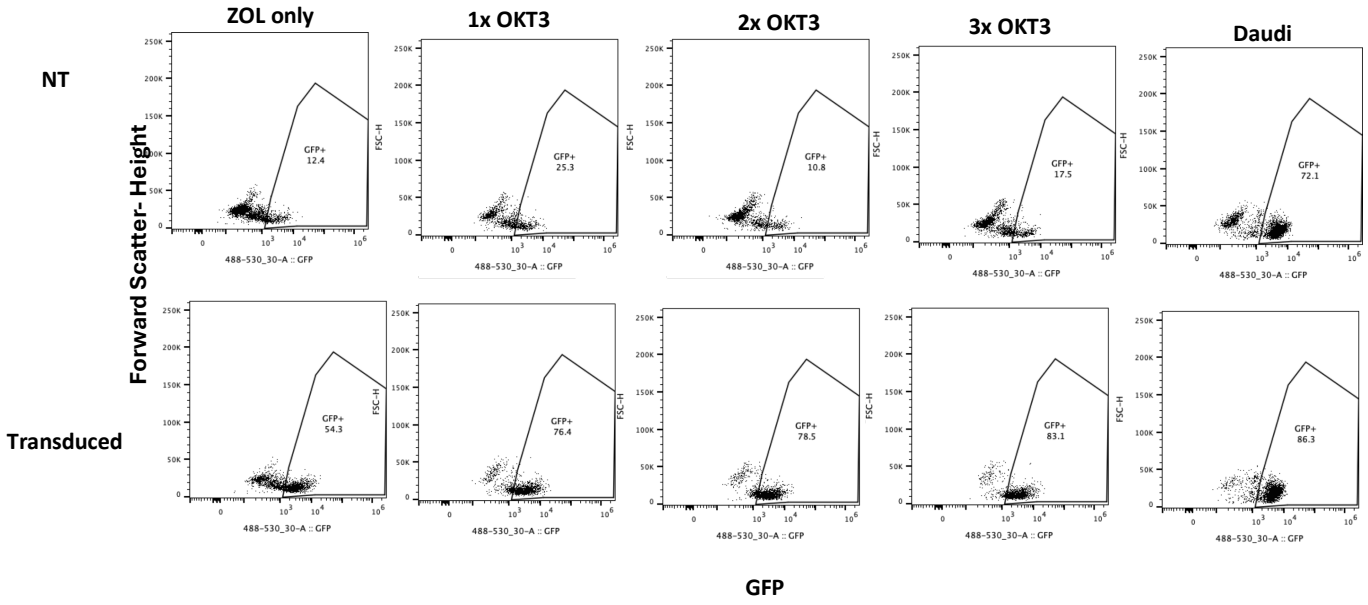


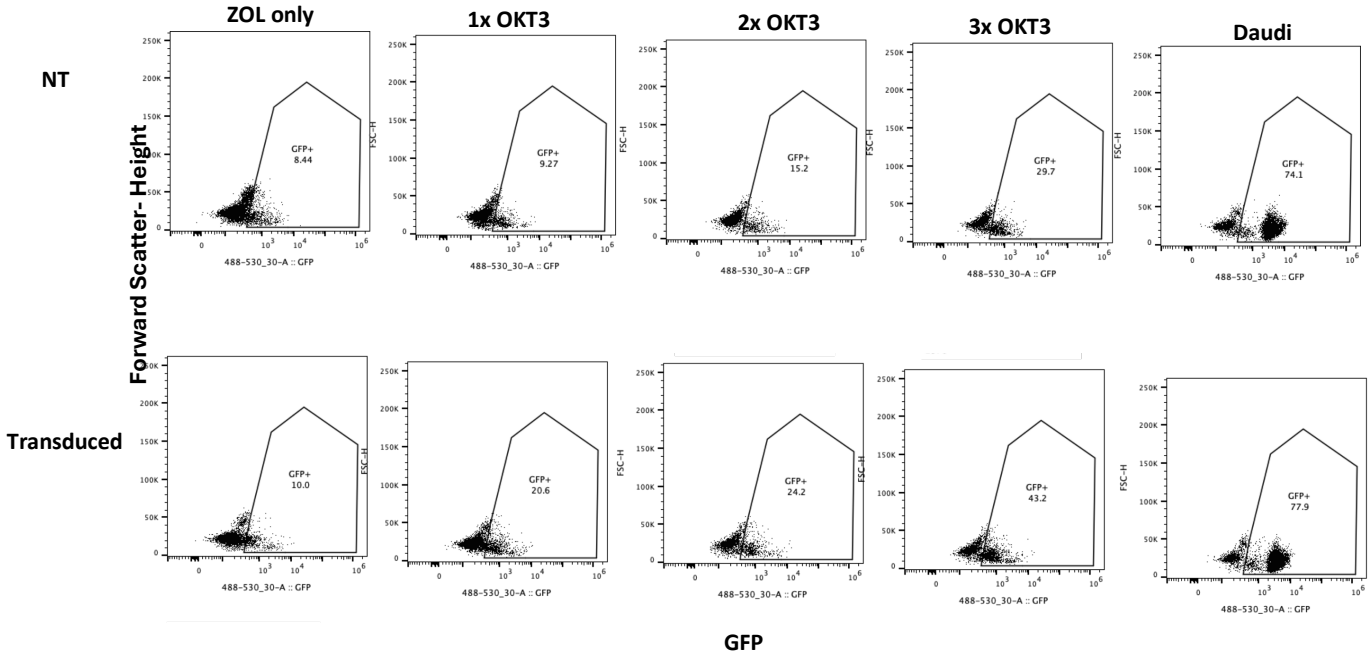

*Figure 6.10: FACS analysis of Vδ2 cells on day 12 (before differential stimulations) for three donors. Transductions are shown for each donor with GFP% cells in addition to the proportion of Vδ2 T cells of both non-transduced controls and transduced Vδ2 cells. NT is used for non-transduced.*

## FACS on Day 19 (T19)

## Donor 1 (LP-2)




## Donor 2 (BC-E)



**Donor 3 (BC-D)**

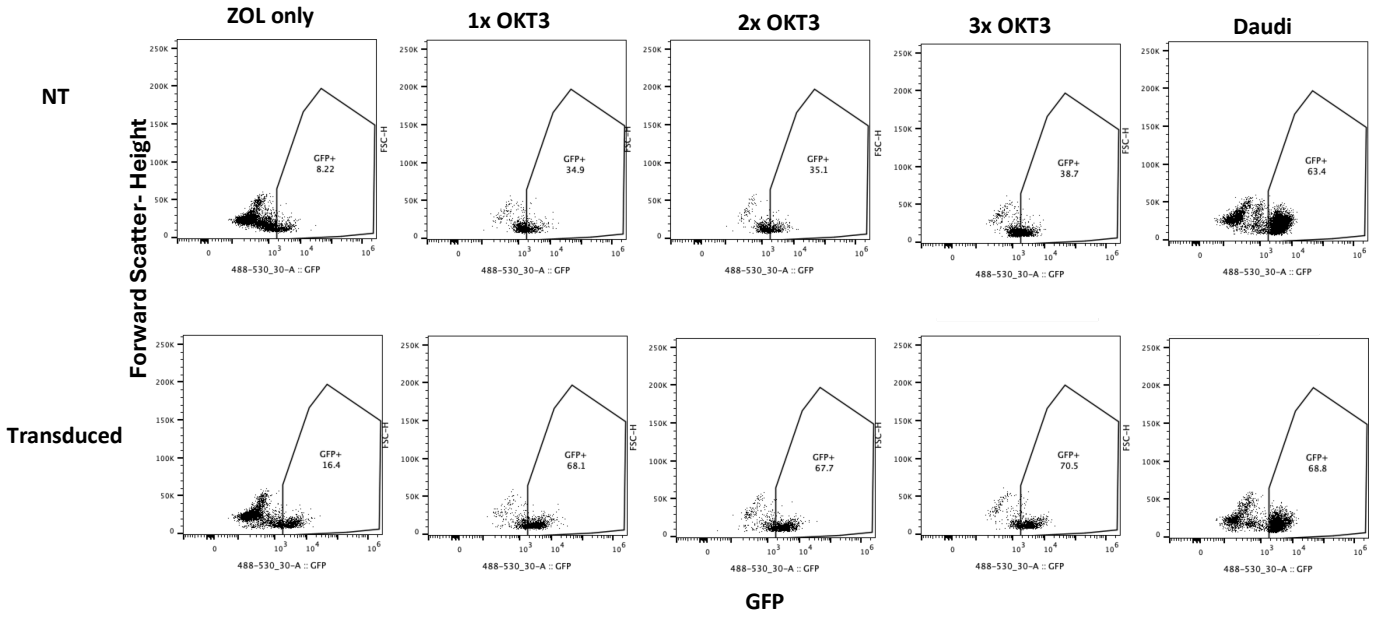



*Figure 6.11: FACS analysis of Vδ2 cells on day 19 (after differential stimulations) for three donors. Transductions are shown for each donor with GFP% of Vδ2 T cells of both non-transduced controls (Top) and transduced Vδ2 cells (Bottom). NT is used for non-transduced.*

In Figure 6.12, GFP percentages are shown for Day 12 post-ZOL stimulation for the three donors, and it clearly represents the lack of good transduction in Donor 2 (BC-E) compared to Donor 1 (LP-2) and Donor 3 (BC-D). The figure also shows the differences in GFP enrichment for the CRISPR library on day 19 (sample collection day) between the different conditions in each donor. Daudi conditions resulted in poor transductions as it had the lowest GFP enrichments. Moreover, as explained previously, it shows clear differences in GFP enrichments between all different donors, with Donor 1 (LP-2) as the highest enriched and Donor 2 (BC-E) as the lowest enriched donor.

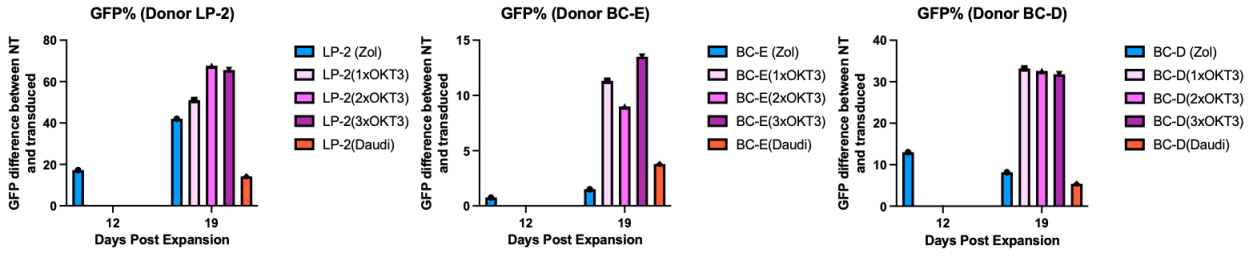



*Figure 6.12: GFP differences of Vδ2 T cells between non-transduced and transduced analysed on day 12 (from ZOL expansion) and day 19 (including all stimulation condition). GFP enrichments on sample collection day (day 19) are shown per donor for each condition.*

*Sample preparation for CRISPR screen analysis by NGS*

The samples will be sequenced by NGS for their gRNA representation at day 19. From the cloned library QC data, an assumption of uniform distribution of gRNAs is present in the viral vector and hence, the T cell sequencing data will be from day 19 only for all conditions and donors.

The cell pellets collected on day 19 were DNA extracted. Then, they were amplified using PCR to generate amplicons containing gRNA cloning sites to send for unique sequence identification using next-generation sequencing. Figure 6.14 shows the location of gRNA library between a U6 promoter (widely used for the expression of CRISPR gRNAs) and gRNA scaffold (important for binding to Cas9 enzyme), flank 1 and flank 2 are regions before and after the gRNA library site which align in part to either the U6 promoter (for Flank 1) or the gRNA scaffold (for Flank 2). The forward and reverse primers are also shown in relation to the gRNA library site. The PCR products were generated for all day 19 samples and run on 1% agarose gel to assess the PCR performance. From all the samples, a band of the correct size (500 base pairs) was obtained and is shown in Figure 6.13. The sequencing method was performed using pair-end sequencing with 2x250bp bi-directional sequencing of the 500bp fragments containing the gRNA library cloned sequences approximately in the middle of the fragments. In detail, the sequencing was done by generating 2 fragments with a length of approximately 250bp from each primer. The forward primer was designed with more than 250bp from the gRNA library sequences (shown in figure 6.14) and is unlikely to be read by the 250bp sequencing, which might cause a lot of gRNA hits to be missed. For the analysis pipeline the 2 fragments generated are first trimmed for adapter reads (to exclude adapters from the generated sequences) and then merged by overlapping them in the central regions of overlap. Finally, the forward and reverse primer sequences (the primers used for PCR amplification) are used to find the correct frame for the unique sequence identification. Any sequences with different length or containing 1 different base pair from other sequences generated is identified as a unique sequence and provided as a table with all reads and counts per read. The unique sequences tables are then filtered by looking for regions before or after the gRNA sequence location (shown in Figure 6.14 with gRNA and depicted as Flank sequences 1 and 2). Sequences that are found to be adjacent to any of the flanking regions are interrogated or identity of the gRNA sequences to be identified. All gRNA sequences identified for each condition are shown in Figure 6.15. The gRNA sequences are also present at different

counts to reflect the enrichment of each gRNA in the samples analysed. The most enriched sequences in each donor condition are shown with blue bars (Figure 6.15). The percentage of sequences with gRNA plus any flanking regions ranges between 0.1 to 6% of the total unique sequences identified and generated from the sequencing run. Details about total unique sequences found with percentages of sequences containing either flanking region (Flank 1 or Flank 2 shown in Figure 6.14) and the number of gRNAs identified per condition are shown in Table 6.1.

## DNA extraction and Amplicon generation and CRISPR-Cas9 KO analysis

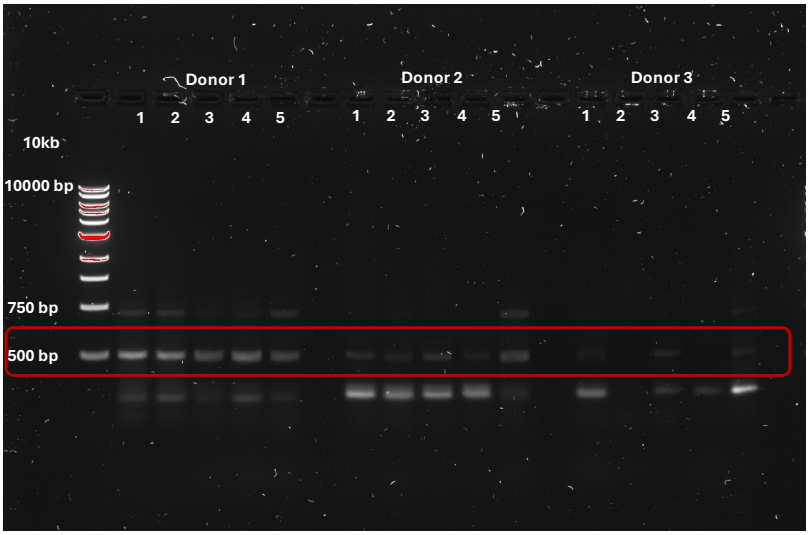



*Figure 6.13: Gel Electrophoresis(1%) showing validation of amplicon PCR prior to sequencing.*

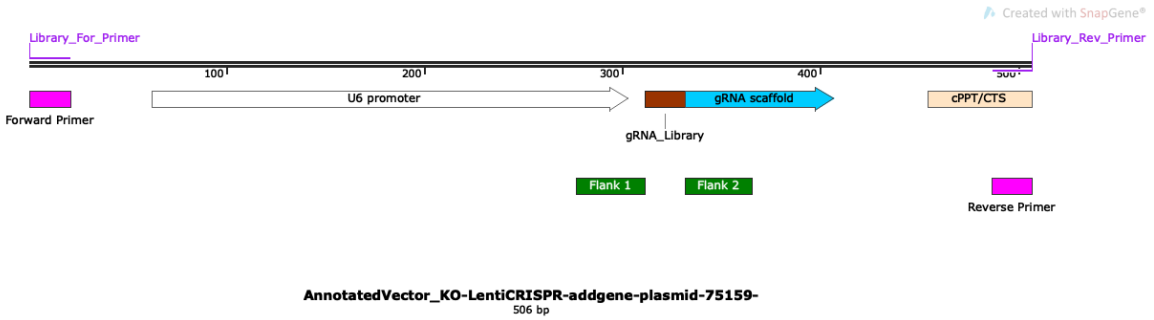



*Figure 6.14: A diagram of the amplicon sequence generated with PCR and sent for CRISPR screen analysis using next-generation sequencing for unique sequence identification. The gRNA library is shown in red, flanked by two green sequences, Flank 1 and Flank 2.*

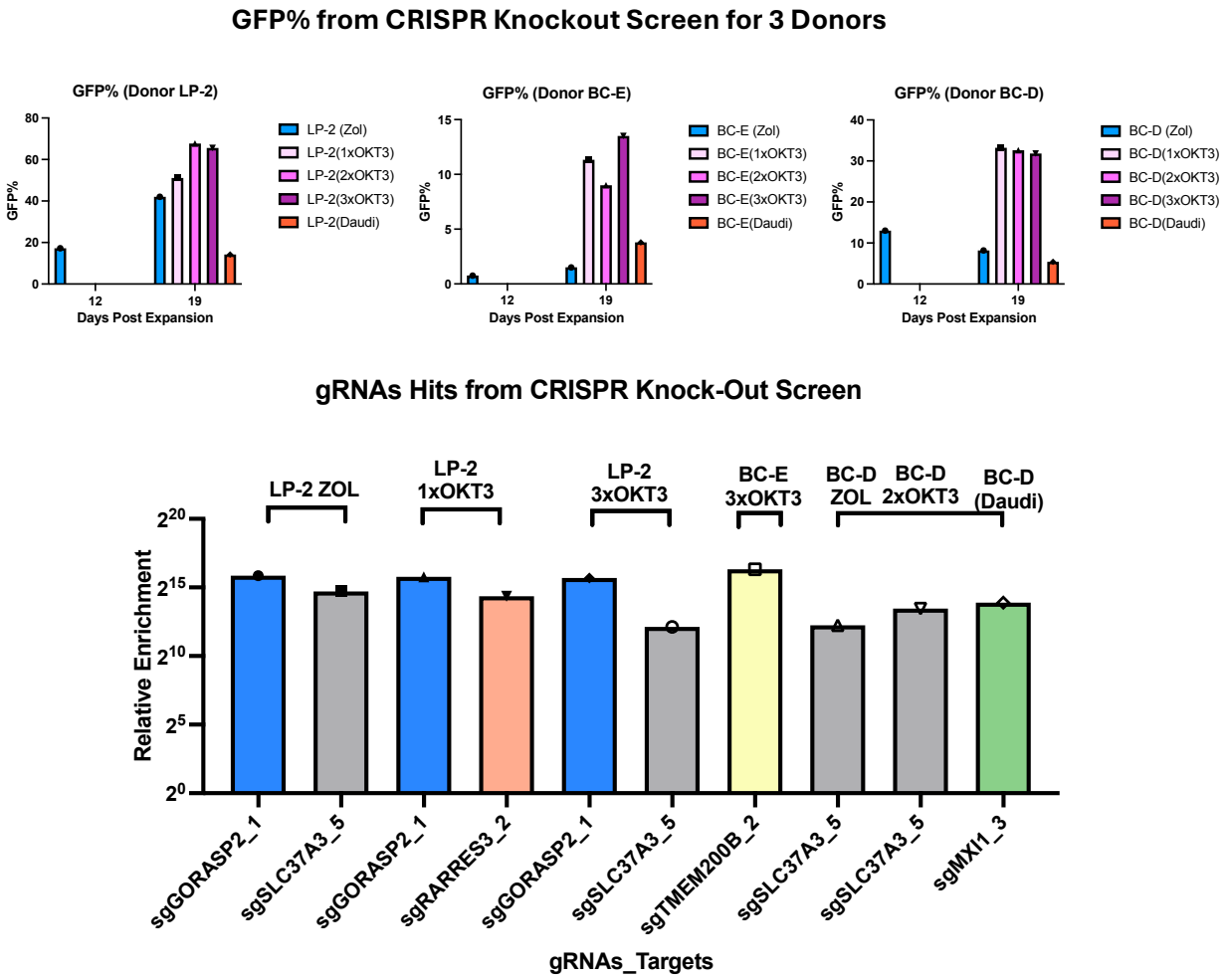



*Figure 6.15: Top: reproduced from figure 6.13 to show GFP% differences in relation to gRNAs hit enrichments in bottom. Bottom: plot of most enriched gRNA targets, similar gRNA are coloured the same while unique gRNAs are coloured differently.*

*Table 6.1: Ratios of identified gRNA targets from total sequencing reads in each donor condition and the percentages of correct sequencing reads including Flank 1 or Flank 2 from amplified DNA.*

| | Total number of unique reads | % of seq with flanks | Number of gRNAs Identified |
|---|---|---|---|
| **LP-2 ZOL** | 9041 | 1.50425838 | 25 |
| **LP2 1OKT3** | 5915 | 0.67624683 | 7 |
| **LP2 2OKT3** | 7460 | 1.17962466 | 20 |
| **LP2 3OKT3** | 5833 | 1.23435625 | 13 |
| **LP2 Daudi** | 12463 | 0.97889754 | 29 |
| **BC-E ZOL** | 5569 | 0.52073981 | 11 |
| **BC-E 1OKT3** | 3859 | 0.33687484 | 4 |
| **BC-E 2OKT3** | 3169 | 6.05869359 | 2 |
| **BC-E 3OKT3** | 4458 | 4.32929565 | 9 |
| **BC-E Daudi** | 17990 | 0.60589216 | 12 |
| **BC-D ZOL** | 10879 | 0.36768085 | 18 |
| **BC-D 1OKT3** | 10395 | 0.12506013 | 7 |
| **BC-D 2OKT3** | 9177 | 0.40318187 | 18 |
| **BC-D 3OKT3** | 12063 | 0.10776755 | 8 |
| **BC-D Daudi** | 22236 | 0.77352042 | 19 |

Following the final analysis of gRNA enrichments, as seen in Figure 6.15, for all donors, there seems to be a preferential expansion with ZOL-only, as only a few of the gRNAs are present from the full library where the majority of gRNAs are not present in all conditions, indicating that absent gRNAs failed to persist with Vδ2 expansions.

In donor LP-2, there was an enrichment of GFP percentage in the ZOL-only conditions, whilst this enrichment was not seen in the other donors. This correlates with the fact that 2 guides were significantly enriched in the ZOL-only conditions, whilst the other donors showed no significant guide enrichment in ZOL-alone conditions.

In donor LP-2, one of the guides enriched in ZOL-only conditions (GORASP2) was also enriched in the 1x and 3x OKT3 conditions, increasing confidence that it is a real hit. Also, in LP-2, one of the ZOL-only enriched guides (SLC37A3) is upregulated in 3x OKT3.

In donors BC-E and BC-D, there is a different pattern of enrichment since ZOL-only does not enrich GFP, Whilst the OKT3 do lead to GFP enrichment. Therefore, you would predict for these donors that you would not see significant guide enrichment in any of the ZOL sequencing but would see it in the OKT3. This turns out to be true. There are 5 guides significantly enriched in these 2 donors, and one of them, SLC37A3, is in common with donor LP-2.

Overall, there is only 1 guide that is enriched in two donors in different conditions, which is SLC37A3 and is the hit of greatest interest. GORASP2 is significantly enriched in only one donor but in three independent conditions.

Daudi conditions were associated with the lowest GFP cells for all donors. They were not shown to have shared gRNA enrichments with any other conditions of the same donor or between different donors. Moreover, no enrichment in sequencing read counts was shown with Daudi-stimulated samples. Daudi conditions seem inadequately transduced and have high background GFP fluorescence resulting in least success of transductions and no real gRNA hits from the CRISPR screen in Vδ2 T cells.

All gRNAs hits are shown (Figure 6.15), similar gRNAs targets between different donor conditions are coloured the same. gRNAs hits were classified as hits with more than 8-10 read counts in each condition, all other gRNAs with reads less than 8 are considered non-enriched and therefore are not shown. Relative enrichment of hit gRNAs was measured as the total number of reads per gRNA over the average number of reads per gRNA. The average number of reads per gRNA was calculated as the total number of gRNAs over the total number of all reads in a sample. The enrichment is shown for each hit gRNAs in Figure 6.15. GFP percentages for each donor are also shown in Figure 6.15 to correlate the levels of GFP enrichment to gRNAs enrichments. With Zol condition, high GFP enrichment is only seen with the donor LP-2 and hence it is the only donor with Zol condition to show significant gRNA hit counts and is the only one shown in the gRNA hit Figure 6.15, as donors with no significant hits of more than 8 counts are excluded. The figure shows a recurring gRNA hit of SLC37A3 gene between 2 different donors ZOL and OKT3 conditions (LP-2 ZOL, LP-2 3xOKT3, BC-D ZOL and BC-D 2xOKT3). Donor LP-2 had a preferential enrichment of gRNA targeted to GORASP2 gene and was with

LP-2 ZOL, LP-2 1xOKT3 and LP-2 3xOKT3 conditions. LP-2 1xOKT3 had a specific enrichment for RARRES3 gRNA, BC-E 3xOKT3 had a large enrichment of TMEM200B and BC-D Daudi had MXI1 targeted gRNA enriched. In summary, there were 2 interesting enrichments observed from the CRISPR screen analysis and that are shared between two different donors (SLC37A3) or is exclusively enriched in one donor but for multiple conditions (GORASP2). Further analysis of these two hits will be performed to understand if they have any important roles in T cells biology.

### *gRNAs hits*

SLC37A3 (Solute carrier family 37 member 3) is localised in the endoplasmic reticulum and lysosome membrane. The expression of SLC37A3 at RNA level can vary in human tissues but few sites include muscle tissues, lymphoid tissues, connective and endocrine tissues (Human Protein Atlas, 2024). It is known to function with the transport of nitrogen-containing-bisphosphonates (N-BPs). In detail, it co-interacts with another protein called ATRAID (All-Trans Retinoic Acid-Induced Differentiation Factor) to form a transport complex for the N-BP molecules that are trafficked to the lysosomes to be released to the cytosol (Yu et al., 2018) SLC37A3 was also reported for its expression in Hepatocellular carcinoma (HCC). SLC37A3 was found to be significantly increased in HCC tissues. Patients with high levels of SLC37A3 in HCC had reduced survival rates compared to patients with low SLC37A3 expression in their HCC cells. knockdown of SLC37A3 in vivo inhibited tumorigenesis and suggested that it has a role in regulating glucose homeostasis (Meng et al., 2023).

GORASP 2 (Golgi reassembly-stacking protein 2) is localised in the Golgi apparatus and re-localises to the ER membrane during ER stress (Kim et al., 2016; Piao et al., 2017). It has roles as a structural protein and is essential for maintaining the formation of Golgi ribbons in the Golgi apparatus ((Zhang & Seemann, 2020) It also may be involved in the regulation of intracellular transport or presentation of transmembrane proteins (Kuo et al., 2000). GORASP 2 expression is less specific in human tissues but at RNA levels its expression sites include the brain, endocrine

and muscle tissues and the pancreas (Human Protein Atlas, 2024). On a protein level, it is expressed in the respiratory system and gastrointestinal tracts. It is also expressed in lymphoid tissues and the skin at both an RNA and protein levels (Human Protein Atlas, 2024).

RARRES3 (Retinoic acid receptor responder protein 3), also known as PLAAT 4, is known to have acyltransferase activities. It is expressed mainly in the kidney, lymphoid tissues, respiratory tract and gastrointestinal tract at RNA levels (Human Protein Atlas, 2024). RaRRES3 regulates signalling pathways and biological processes, such as tumour cell proliferation. In one study, the effects of RARRES3 overexpression resulted in the inhibition of Wnt/β-catenin signalling activation, which causes a suppression of stemness properties (Hsu et al., 2015).

TMEM200B (Transmembrane protein 200B) is a transmembrane protein with no specific known functions, especially with T cells. It is generally known to be expressed at RNA levels in muscle tissues, endocrine tissues and the gastrointestinal tract (Human Protein Atlas, 2024).

MXI1 (Max-interacting protein 1) is an antagonist for MYC transcriptional activity. It acts as a repressor by competing with MYC to bind to MAX to form a DNA-binding heterodimer, causing transcriptional repression (Zervos et al., 1993). MXI1 can be expressed in the eye and connective tissues at RNA levels and expressed at protein levels in lymphoid tissues, muscle, kidney, Liver, gastrointestinal tract, endocrine tissues and the brain (Human Protein Atlas, 2024).

Little is known about the roles of these proteins in γδ T cells or if they have other roles than intracellular transport of complexes or other molecules. To study their functions with Vδ2 T cells and T cell exhaustion, it is necessary to make individual knockouts of each gene and study the effects on Vδ2 cells at a molecular level.

## Activation of target genes using CRISPR dCasmini system

Several genes (18 targets) from the RNA sequencing analysis associated with significant differential expression in OKT3 samples were selected. The dCasmini system for activation will be used for overexpression of the selected genes in Table 6.2. Few genes were selected from publications of exhaustion in ab T cells and were shown for their effects on enhancing T cell's functionality and reversing exhaustion. Each gene was targeted by four gRNAs designed by CRISPick software.

*Table 6.2:List of genes selected for activation in Vδ2 T cells.*

| Gene targets for activation library | Number of gRNAs |
|---|---|
| **ANXA1** | 1 |
| **C1QTNF3** | 3 |
| **CALCOCO1** | 2 |
| **CLEC2B** | 3 |
| **CX3CR1** | 4 |
| **FAM13A** | 4 |
| **GPR155** | 1 |
| **IL18RAP** | 3 |
| **JUN** | 2 |
| **PLAC8** | 3 |
| **RARRES3** | 2 |
| **S1PR4** | 3 |
| **SAMD3** | 4 |
| **SH3D21** | 3 |
| **SKAP1** | 3 |
| **SLAMF6** | 4 |
| **T-bet** | 1 |
| **TCF7** | 3 |

### *dCasmini-VPR virus titrations*

The dCasmini system was used to clone an activation library using a gRNA pool with selected genes and number of gRNAs cloned per gene are shown in Table 6.2. In addition to the gRNAs library vector, a deactivated Casmini enzyme that is fused to a transcriptional activator

(dCasmini) is used. The main objective of the library is to screen for the effects of overexpression of target genes in Vδ2 T cells.
As the Casmini system is a two-vector system, virus production and titration were made separately for the dCasmini activator enzyme (containing mcherry in the backbone) and the bespoke gRNA oligo pool. The gRNA pool was ordered as a synthesized oligo pool from IDT and then cloned in a vector backbone (Xu et al., 2021) The backbone contains a blue fluorescent protein (BFP) to be detected in FACS. The cloned gRNA pool was assessed by sending the DNA for next-generation sequencing after amplifying the region containing the gRNAs. The sequences were then filtered to check for the gRNAs cloned and ensure how many gRNAs per guide were successfully incorporated in the library. The number of gRNAs cloned per gene is shown in Table 6.2.

Following the success of the cloning, Virus production was started, and the RDpro envelope virus was made for both the dCasmini vector and gRNA pool vector in a similar way to the CRISPR knockout library. The virus was then tested for titration on Lenti X 293T cell lines.

Virus titrations showed high efficiencies in transducing lenti X 293T cells with the gRNA pool (Figure 6.16). The titrations showed 72% BFP-positive cells using 10ul of the virus with 100,000 cells. Virus volumes of 2ul, 0.4ul and 0.08ul associated with 34.6, 11.6 and 7.9 BFP% cells. Successful transductions were encouraging for the following tests on γδ T cells. However, the analysis of the dCasmini vector led to very low transductions (Figure 6.17). Repetition of virus titrations always led to similar results where the gRNA oligo pool corresponds to high efficiency and ease of transduction with low volumes but the mcherry dCasmini enzyme never resulted in significant mcherry positive cells following transductions. The screen of the dCasmini and activation gRNAs pool was not tested further due to the dCasmini vector issue and lack of transductions.

## gRNA Oligo Pool titrations on Lenti X 293T cells

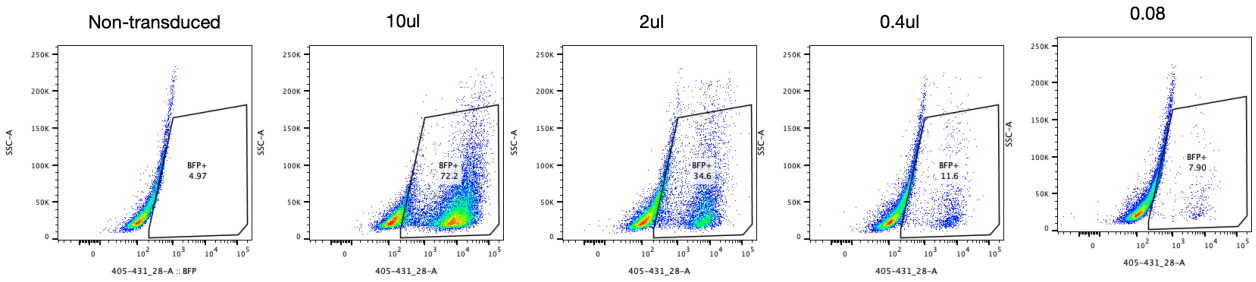



*Figure 6.16: Virus titrations of gRNA pool vector on Lenti x 293T cells lines. Four different virus volumes are used and shown in comparison to the non-transduced control.*

## dCasmini titrations on Lenti X 293T cells

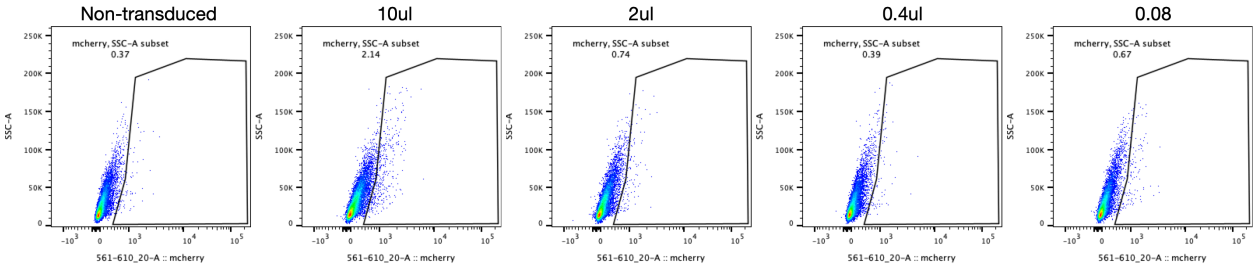



*Figure 6.17: Virus titrations of dCasmini vector on Lenti x 293T cells lines. 4 different virus volumes are used and shown in comparison to the non-transduced control.*

## Conclusion

In conclusion, this chapter evaluated the use of different factors to produce a high-titre CRISPR lentivirus for Vδ2 T cells. A protocol was optimised with a new envelope (RDpro) for virus packaging with the transfection agent GeneJuice to produce the virus. Then, retronectin and spinfection were applied to transduce Vδ2 T cells effectively.

A pilot study was performed to test the ability of the transduced Vδ2 T cells to express GFP over 19 days and was successful. Based on the pilot study, a final CRISPR screen assay was developed to test the CRISPR knockout library of Vδ2 T cells over 19 days, using differential stimulation with OKT3 and including a Daudi stimulus. The CRISPR screen analysis identified a small number of gRNA hits for genes that were significantly differentially expressed in the RNA seq for the ZOL and OKT3 conditions. gRNA hits were defined as gRNAs with 8 or more reads and thus with calculated relative enrichment of 4400 or higher. Guides that had enrichments in more than one donor added confidence of being real gRNA hits. Two gRNAs had high enrichments; one was for the gene SLC37A3 which was present in multiple donors (LP-2 and BC-D) with ZOL and OKT3 stimulations; two was for the gene GORASP2 which was present in 1 donor but with high counts and between different conditions (ZOL and multiple OKT3 conditions). The two identified gRNA hits are targeted to genes from the RNA profiling and with significant differential expression in OKT3 samples (Figure 6.18).

The gRNA hits did not have established roles in the literature in T cells or specifically in Vδ2 T cells. However, all significantly enriched gRNA hits need to be further investigated and validated with a single knockout assay to assess their effects on the functionality of Vδ2 T cells.

There were a few limitations in the CRISPR screen performed. First is the lack of availability of sequencing data of the gRNA pool cloned by TWIST to ensure the representation following the expansion of the plasmid library in bacterial cultured and DNA extraction. The evidence provided by TWIST bioscience was used to confirm the right distribution of all gRNAs in the library. Second, it was not possible to sequence the cells following transducing with the pooled library and prior to adding differential stimulations to get an overview of gRNAs transduced in Vδ2 cells, which was mainly due to the lack of enough cell numbers to add as a none-stimulated

transduced control. Third, the forward primer was designed with more than 250bp from the gRNA library sequences and is unlikely to read through to the guide sequences using the chosen sequencing method (250bp bi-directional sequencing) leading to a potential loss of many gRNA hits from the reads. Fourth, the method used to calculate the enrichment scores of each gRNA was not a standard method and was done with simple calculations to show relative enrichment for each of the gRNA hits found. Benefiting from external bioinformatics expertise could have led to a better representation of the gRNA hits present and increased confidence in the analysis.

Finally, the CRISPR activation dual vector system failed to be tested in Vδ2 T cells due to the inability to transduce the Vδ2 T cells with the original Vector with dCasmini enzyme. Still, the cloned gRNA pool vector was successfully tested and is very efficient in transducing Vδ2 T cells. For the future, it is important to prioritise using a different activation enzyme or further optimisations of the current vector before using the gRNA pool and screens in Vδ2 T cells.

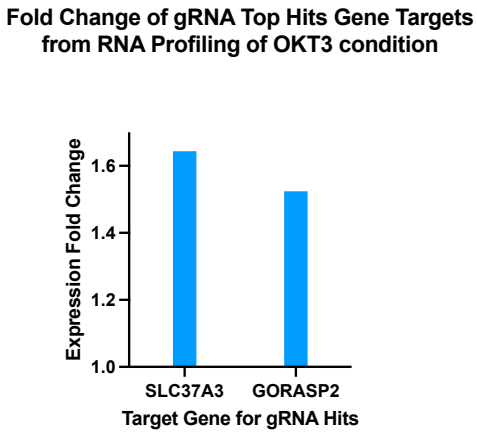



*Figure 6.18: Fold change expression of gene targets of identified gRNAs top hits in OKT3 condition from previous RNA profiling.*

# Chapter 7

## Discussion and Conclusions

### *Chapter 3*

Reduction in Vδ2 cell numbers was used as a potential phenotype of Vδ2 exhaustion that can evaluated in subsequent screening experiments to identify exhaustion features of Vδ2 cells biologically and transcriptionally. Evaluating different models of chronic stimulation, we identified the ZOL model, where the cells get initially stimulated with Zoledronate and followed by a re-stimulation with OKT3, to show reduced Vδ2 cell numbers following the repeat stimulations consistently and is therefore potentially consistent with induction of exhaustion in Vδ2 T cells.

In the other model with different numbers of OKT3 stimulations there was no significant difference in T cell numbers after 15 days expansion when comparing single, double or triple simulation. This OKT3 model was chosen as a standard activation model of T cells that are commonly used in reports or in the clinic. The OKT3 model is well suited for conventional T cell stimulation via the CD3 component of the T cell receptor and has been chosen to be used at a high concentration (1ug/ml) to ensure chronic stimulation of the T cell population during a short period of time. The timeline of stimulation was also chosen from recent work on exhaustion induction in t cells and it is elicited early on from day 6, reflecting the phenotyping of the in vitro model on days 0,3,6,9 and 15.

However, the model did not result in upregulation in the expression of markers of activation and exhaustion, such as LAG-3 and TIM-3, which we did not expect. Limitations of the OKT3 model include the evaluation of only one concentration of OKT3 to stimulate T cells. Moreover, it was not possible to include a study or proliferative potential of T cells in this model. The cells were also predominately of αβ T cell type in the culture; however, the TCR levels were impacted by the repeated stimulation, which could have affected the gating of the full population of αβ T cells in the culture and resulted in misleading information on their state of markers upregulation or their true numbers in the culture. Using a method to reduce or inhibit TCR internalisation

prior to analysis of the cells could increase the accuracy of detection and conclusions made from this specific model, although there is no known standard or validated experimental approach for this. As this model was favourable for αβ T cell expansion, a different model was developed using Zoledronate, a common Vδ2 T cell activator clinically, initially to enrich Vδ2 T cell populations followed by OKT3 stimulations.

This differences in this model from the initial zoledronate stimulation implies that the nature of the initial stimulation (zoledronate or anti-CD3) induces different sensitivity to subsequent OKT3 challenge. There are several possible explanations.

OKT3 antibodies might give stronger and more ongoing initial stimulation than zoledronate'; for example the OKT3 antibodies might persist until the time of the second stimulation so that the cells are already maximally stimulated in terms of TCR ligation. Therefore, addition of the second antibody has no additional effect on T cell signaling. In contrast, zoledronate is known to be dependent on antigen presenting cells to engage TCR through butyrophilin 3A1 and 2A1 . It is possible that the the ZOL stimulation is much more short lived either because the APC (antigen presenting cells) disappear from culture conditions or the effect of zoledronate is short lived. It is also possible that the strength of TCR signal from Zoledronate mediated TCR ligation is weaker or qualitatively different to that imparted by OKT3.

The timeline of the Zol was also shortened to focus on investigating exhaustion elicited on day 6, as recommended by the literature. A proliferation study and novel exhaustion marker TOX and progenitor T cells marker of stemness TCF-1 were included, which is common on the pool of T cells before the development of terminal exhaustion. A significant decrease in cell number was seen with the multiple stimulated conditions from the ZOL model and associated with significantly higher activation and exhaustion markers (TIM-3, LAG-3 and TOX) expression. It also showed interesting TOX-TCF-1 dynamics in Vγ9Vδ2 T cells. Increased levels of TOX indicate the existence of cells that are exhausted or initiating an exhaustion program (Khan et al., 2019), and with TCF-1 expression increasing with repeated stimulation, it possibly means the existence of a heterogeneous population of Vδ2 cells following OKT3 stimulation with cells that are of an exhaustion phenotype and some of an exhaustion progenitor cells characteristics (Beltra

et al., 2020; Blank et al., 2019; Mann & Kaech, 2019). Moreover, the proliferation dilution assay identified a potential emergence of an exhausted population following the second stimulation (ZOL + 1xOKT3 condition) that also expresses significantly higher levels of markers associated with exhaustion in T cells (TIM-3, LAG-3, TOX and TCF-1), unlike TOX and TCF-1 that have been known to be associated with different exhaustion stages in T cells, TIM-3 and LAG-3 are regarded as activation markers of T cells in addition to being exhaustion markers which make the evaluated data regarding those two markers inconclusive for exhaustion. However, with the expression patterns seen with TOX and TCF-1, there is increased confidence that TIM-3 and LAG-3 expression could represent a phenotype of exhaustion or exhaustion induction (Anderson et al., 2016). Nevertheless, it is crucial to remember that the decrease in numbers might at least in part be due to activation-induced cell death by day 9; however, there is no clear evidence in terms of data regarding that phenomenon, making it harder to make conclusions about the developmental state of the population as a whole.

With the Zol model, even though the expression patterns of selected markers are as expected from the literature on exhausted T cells. it would have been beneficial to prolong the model to evaluate what differences could emerge in terms of the cell counts and marker expression. Including markers to study activation-induced cell death, like Propidium Iodide or annexin-V, would have helped in making firmer conclusions. The proliferation study was associated with interesting observations, but it would have been improved if proliferation detection was done every 12 hours instead of every 3 days. The final model was chosen to mimic natural stimulation to Zol via target cells.

The final model was chosen to mimic natural stimulation to Zol via target cells. the model used Daudi as they are often used in the literature for Vδ2 T cell stimulation. The Daudi cells were pre-coated with the 20.1 monoclonal antibody, which forces the interaction between Daudi and Vδ2 T cells via the BTN3A1 and BTN2A1 heterodimer on Daudi which is thought to function as the most relevant ligand for the Vγ9Vδ2 TCR. The results were not as expected of an exhaustion model; still, the model showed some heterogeneity with the proliferation of multiple stimulated samples compared to single stimulated samples. An improvement of this model would be the use of Daudi cells that are not pre-coated with 20.1 antibody to compare with the model used.

In conclusion, exhaustion is difficult to define in this context. However, reduced cell numbers and upregulation of exhaustion marker expression provide sufficient evidence to support proceeding with further evaluation of the ZOL model to potentially identify new biological and genetic insights.

*Chapter 4*

The two main hypotheses in this chapter are: first, an emergence of a population that is viable but hypofunctional and therefore hypo-proliferating (i.e. exhausted) following the double stimulation and second, loss of cells that is caused by cell death after the repeated stimulation. To investigate both hypotheses, cell proliferation by cell trace dilution was tested in the presence or absence of tumour target cells and looking at the cytotoxic function of Vδ2 cells in the cells emerging following the two expansion protocols described.

The first experiment tested the emergence of a hypofunctional hypo-proliferating population. The proliferation and survival of Vδ2 T cells were evaluated in response to stimulation with Daudi target cells at different effector-to-target ratios (24 or 48 hours after OKT3 stimulation); Daudi cells were chosen as they are a proven target for Vδ2 cell stimulation as they accumulate high levels of IPP (Cano et al., 2021). The Persistence and expansion of Vδ2 T cells did not show any significant differences between single or double stimulus with increasing ratios of Daudi target cells (1:1,1:5 and 1:10, Effector-to-target) in terms of cell counts or proliferation. Proliferation responses were similar between different conditions for both Daudi and Vδ2 T cells, suggesting no differences in hypofunctionality of Vδ2 T cells at the chosen time points of the co-culture experimental set-up (24 or 48 hours after OKT3 stimulation). There was no clear evidence of hypofunctionality following stimulations with OKT3 and using the proposed experimental timeline and target cell ratios.

The co-culture with Daudi cells did not provide what is expected of exhausted Vδ2 T cells. The main limitation of this experiment is the short duration of the co-culture after adding the differentiation of Daudi cells. The reason for selecting the current time points is to look at the

differences in responses with cell counts and proliferation before the cells are fully committed to exhaustion or die due to activation induced cells death

Second, the emergence of hypofunctionality and heterogeneity of Vδ2 T cell populations following the dual stimulation with OKT3 was studied by monitoring cell proliferation. The study suggested the presence of heterogenic populations that either include terminal effector cells, preferentially stimulated cells to ZOL or OKT3 that led to their selective proliferation at different stages of stimulation, or cells that experienced activation-induced cell death and therefore disappeared from the culture and are not shown in snapshots taken at days 4 or 5. A small population of exhausted Vδ2 T cells might not have proliferated or died and thus disappeared from the proliferation assay, causing difficulties in drawing firm conclusions regarding the mechanism of reduced Vδ2 T cells following double or triple stimulations in this assay.

For these studies we chose to evaluate proliferation using dilution of cell trace dye. This technique provides a “historical” insight into proliferation rather than an “in time” assessment. In other words, the degree of dilution informs the observer the cumulative number of cell divisions that have occurred between the time of cell labelling and time of flow cytometric analysis. It does not however give any insight into whether cells are still proliferating at the time of endpoint analysis. For example, a sample which shows high proliferation in terms of cell divisions over a 5-day period could be highly exhausted and non-proliferating at the time of analysis as a result of high activity during the previous 5 days. Alternate techniques are available for determining whether cells are actively proliferating at time of analysis. For example, Ki67 staining is a commonly used marker considered to be specific for active proliferation. Similarly, BrdU incorporation over a time period of 2-4 hours gives insight into the proliferation state during that shorter time period. Moreover, actively dying cells can be distinguished in flow cytometery using techniques such as propidium iodide and annexin-V staining.

Further insights into active proliferation status following second and third OKT3 simulations in the zoledronate model might therefore be obtained by these additional flow cytometry stainings.

Third, testing of differences in response to tumour cell targets for single and double-stimulated cells was achieved. the response of Vδ2 T cells following double stimulations was evaluated by cell counts and expression of exhaustion markers (PD-1 and LAG-3). The hypothesis is that with double stimulations, the cells contain a higher proportion of exhausted cells that will further manifest as reduced cell proliferation and cytotoxicity compared to the single ZOL-only stimulation. The experiment was set with a re-challenge of Vδ2 T cells using different targets (i.e. Daudi, THP-1 and Jurkat) on alternating days. A Re-stimulation at an early stage (1 day after OKT3) did not result in convincing differences in responses from the different conditions. The experiment suggested that early re-stim at the chosen time points does not disclose hypoproliferation characteristics of any of the conditions tested.

T cells were re-stimulated using three different types of target cells. However, the study revealed non meaningful results and it did not correspond to what was expected regarding the different sensitivities of Vδ2 T cells against each of the targets. Moreover, the re-stimulation was done every 24 hours by adding the exact number of cells used at the first stimulation time point. The complexity of calculations at the time of adding target cells could have led to the accumulation of target cells in wells and might not even provided the desired stimulation to Vδ2 T cells. All the analysis was done on day 7, which provided a disadvantage of proper comparison between 1x,2x and 3x stimulations as they were added on day 4, day 5 and day 6, respectively. In detail, the analysis performed was at different durations following the stimulations, 3 days after 1x stimulus, 2 days after 2x stimulus and 1 day after 3x stimulus.

Overall, there were no clear differences at the timepoints evaluated between the two conditions tested in terms of proliferation in the presence or absence of tumor target cells. The two main hypotheses for a reduction in Vδ2 cell numbers following repeated stimulation were one, the emergence of a viable population that is hypofunctional and hypo-proliferating (i.e. exhausted); two, cell loss by death.

The data failed to show hypo-proliferation or reduced cytotoxicity and is more supportive of the hypothesis of cell loss by activation-induced cell death. Still, none of the hypotheses was firmly established or excluded, and there is a possibility that both are true in that the reduction in cell numbers results from both mechanisms suggested.

Evidence supporting exhaustion is the expression of exhaustion markers shown in chapter 3 for the double-stimulated cells, In addition, the lower number of VD2 cells was associated with the double stimulations following the encounter with Daudi at a 1:1 ratio (Day 4 data). Proofs in favour of activation-induced cell death are one, the lack of any evidence for hypo-proliferation or reduced functional cytotoxicity in chapter 4, two, even though is not statistically significant, are the small differences in Daudi cytotoxicity shown at a 1:1 ratio where Daudi numbers are lower with double stimulation samples, however, this favours surviving cells having enhanced cytotoxicity instead of reduced after the double stimulation.

Further research could provide promising insights to test these hypotheses. For instance, performing stains for cell death markers at regular intervals, such as every 12 hours following the second OKT3, could provide valuable information. Additionally, evaluating cytotoxicity at higher E:T ratios, under more stressful conditions, or against a less proliferative target could also yield significant findings.

While we cannot definitively conclude whether T cell exhaustion or AICD is the more likely cause of the reduced Vd2 cell number following double stimulation, the data has provided sufficient evidence of functional differences between the two stimulation conditions. This provides an opportunity to evaluate whether double stimulation causes a transcriptional change, an interesting area for further investigation.

*Chapter 5*

From the ORA analysis, multiple gene sets (immune regulation, T cell differentiation, and responses to viruses) indicate that immune response has been triggered from OKT3 stimulation. Also, cnet networks showed genes associated with these sets, like BATF3 and JunB, further

suggesting differentiation responses with double stimulation. Some sets prove the presence of chronic stimulation and increased cellular stress that trigger responses like cytokine receptor, signalling and DNA repair.

The significant differential expression of BATF3, LAG3, TOX2, NR4A, TCF4, TCF12 and TCF7L2 in the double-stimulated γδ T cells is consistent with exhaustion and exhaustion progenitor pathways on day 6 of expansion, three days after OKT3 stimulation.

BATF3 expression was significant in both RNA profiling experiments and might indicate induction of specific pathways through BATF3 and AP-1 family proteins responsible for T cell differentiation, such as JUN, JUNB and FOS. BATF3 is an AP-1 family inhibitory protein known to have increased expression in exhausted T cells, where it counters the expression of JUN and FOS of the AP-1 family and consequently reduces T cell functionality (Lynn et al., 2019).

Few of overexpressed genes were from apoptotic and anti-apoptotic pathways, which agree with previous findings and the suggestion that there might be a few cells that are dying following OKT3 stimulation, causing the reduction in cell numbers (Figure 8, chapter 3) and the inability to be detected with functional assays such proliferation by dye dilution. Some significant differentially expressed genes promote cell survival and proliferation, ensuring cell stress and activation were triggered with the OKT3 model.

Exhaustion-related gene sets showed strong evidence of similar gene signatures between the double-stimulated γδ T cells and exhausted CD8 T cells. One study (GSE9650) was significantly enriched with both RNA profiling data sets (Wherry et al., 2007), confirming the *in vitro* model's success in inducing chronic stimulation and suggesting the initiation or development of T cell exhaustion in the double-stimulated Vγ9Vδ2 T cell. Specific associations of differentially expressed markers to Vγ9Vδ2 T cell exhaustion and development will be confirmed with a CRISPR Knock Out screen of pre-selected targets.

The use of CRISPR screens to investigate the effects of a set of target genes on the functions of T cells has been widely used and very effective. In gamma delta (γδ) T cells, there have not been any reports on using a CRISPR screen to knock out multiple genes in a pooled format. Transductions using CRISPR viruses in γδ T cells have also not been established; therefore, it was required to evaluate multiple methods to ensure the ideal protocol is used to transduce Vδ2 T cells and evaluate the CRISPR screen. Following multiple methods for transducing T cells, it was identified that the ultimate protocol for successful transduction was using multiple factors. First, the lentivirus is packaged with a RdPro envelope using Gene juice as the Lenti X 293T cell transfection agent. Then, Vδ2 T cells are transduced using 100ul virus per 100,000 Vδ2 cells on retronectin pre-coated plates with spinfection. Multiple factors during virus production and transduction time ensure the success of the transactions with Vδ2 T cells, as shown in the pilot study where the cells were GFP positive for the CRISPR library even after 19 days in culture and following OKT3 stimulations. The transduced library was not re-validated following the expansion in bacterial culture to ensure full representation of gRNA targets. This could be a great limitation of the study as we cannot make sure the full library was used in the experiment.

GFP enrichments were clear (Figure 12 Top) and are consistent with gRNA hits found (Figure 12 Bottom), increasing the confidence of enrichments due to gRNAs. However, there is a potential that Cas9 by itself or GFP alone provided the preferred expansion to Vδ2 T cells. Using a control like GFP-Cas9-only vectors to be transduced in parallel would have increased the certainty in the results shown even more. The inclusion of transduced but non-stimulated cells and cells analysed for gRNA representation from Day 0 would have provided a better experimental set-up for the CRISPR screen and hits identification.

The CRISPR screen next-generation sequencing analysis yielded unique reads for each donor and condition. Cells stimulated with Daudi had the largest number of reads in each donor, even though they were associated with the least GFP enrichments. A good reason could be reads resulting from the background plasmid with no gRNAs. Moreover, the unique reads analysis

works by considering all sequences with different lengths or with different single or more basepair variations as unique sequences and thus increasing the total counts of non-useful reads in each sample. This also could explain why the percentage of sequences with flanks is very low for all donors. Another reason is that some of the read lengths are short due to reads that have not reached the flank regions with sequencing and did not cover the region with the gRNA library.

Few gRNA hits were Identified from the CRISPR pooled screen; SLC37A3 is the most interesting as it was enriched with 2 different donors for ZOL and OKT3 conditions. This indicates that SLC37A3 might have a role in Vδ2 T cells and could affect the expansion potential of Vδ2 T cells. GOSRASP2 was enriched in 1 donor but with multiple conditions and could be of interest to investigate further as its high counts are convincing for it to be a real hit. The other gRNAs like RARRES3 and MXI1 are much less convincing and interesting as they do not show enrichments in multiple donors or different conditions.

Validation of the hits of the CRISPR screen is required to conclude if these gRNAs have any roles in Vδ2 T cells and if they do impact the persistence and exhaustion of these cells. A single knockout of each gRNA (SLC37A3 and GORASP2) will be necessary to study the effects individually on the functions of Vδ2 T cells following stimulations with OKT3.

Finally, an attempt was made to evaluate a mini CRISPR activation library but using the vector for transduction in Vδ2 T cells was unsuccessful. The dCasmini vector is a two-vector system with a gRNA library cloned in a second vector to the dCasmini enzyme. The gRNA library was cloned successfully and transduced Vδ2 cell with high efficiency. However, there were issues with the transduction of the main unmodified vector containing the Cas enzyme with the activation domain. For a potential screen of this library in the future, a new vector with a dCasmini enzyme needs to be tested and, upon successful use on Vδ2 T cells, can be combined with the cloned gRNA library vector and screened for potential target overactivation in Vδ2 T Cells.

## Future directions

It might have been more informative to test the effects of multiple concentrations of the OKT3 model on inducing exhaustion. It would have been a great advantage to learn how the cells proliferated over time and to further confirm if the lack of upregulation of TIM-3 and LAG-3 is associated with increased proliferation or lack of activation.

For the stimulation models, it is recommended to perform a titration of the OKT3 antibody and extend the duration of stimulation. The concentration used could not provide optimal conditions for chronic stimulation for the period of stimulation tested. Performing cell counting in addition to classical and exhaustion markers would be more beneficial if it was done side by side with proliferation and death markers and at more frequent analysis time points. The use of non-20.1 coated Daudi and varying the frequency of Daudi addition to Vδ2 T cells could also lead to new insights into exhaustion induction in Vδ2 T cells.

For the Daudi co-cultures, in addition to the times tested, the previous models showed that hypofunctional Vδ2 T cells emerge between day 6 and day 9 based on cell count reduction and upregulation of exhaustion markers. It might have been beneficial to set the co-cultures between day 6 and day 9 in addition to day 4 and day 5, which could teach us more about the populations that we hypothesised to be functional. Also including a read-out 72 hours after stimulation would be informative. To better assess the functionality of Vδ2 T cells detection of proliferation at shorter periods such as every 4 to 12 hours instead of every 24 hours could reveal more about the dynamics of proliferation in the multiple stimulated cells. Moreover, including cell death markers and a few exhaustion markers to detect the heterogenous populations shown with cell trace tracking could provide more exact phenotypes of each population developed following the single and double stimulation over the period tested. Expanding the proliferation study up to day 10 or 12 would be very informative in terms of showing the fate of populations shown as hypoproliferative. The re-stimulation with target cells would be better if the restimulation were more than 24 hours apart and extended to longer than 7 days. The analysis might have been more informative if it were done following 1,2,3, and 4 days of each stimulation. Moreover, varying the ratios used for effector-to-target cells would increase the quality of information learned from such a model.

The transcriptomic experiment showed heterogeneity of the cultures tested and impacted the confidence of the analysis done as it is hypothesised that a high proportion of the cells on day 6 are of exhaustion phenotype. However, a technique such as bulk RNA sequencing does not offer the high resolution of a single-cell sequencing experiment. It is greatly recommended to repeat the experiment with single cell transcriptomics analysis, which has the potential to show different transcription profiles of each single cell and show in greater detail the associated transcriptome of that cell.

The CRISPR screen had some limitations. The library was not sequenced following the expansion in bacterial cultures to ensure the representation is still the same for all gRNAs. Moreover, following transductions of Vδ2 T cells, the cells were not sequenced at an early stage prior to differential stimulation for conditions acquired. Only cells at the final collection time point were analysed. For future experiments, it would be recommended, if possible, to perform a sequencing step following the expansion of the plasmid library in bacterial cultures and before adding differential stimulations to select surviving cells. For analysing the CRISPR screen, the performance of standard methods of CRISPR analysis via bioinformatics tools is desired to ensure validation and identification of hits. Finally, analysis of hits by doing a single knock-out assay for each gRNA hit individually in Vδ2 T cells is required to further assess the importance and role in Vδ2 T cells exhaustion or persistence following chronic stimulation.

## Positive outcomes of the study

- One protocol for repeat Vδ2 cell stimulation (ZOL model) leads to reproducibly reduced Vδ2 cell expansion compared with single stimulated cells
- Repeat stimulation leads to the expression of genes that are associated with exhaustion in αβ T cells
- The study adds to a growing body of literature that exhaustion is a phenomenon in Vδ2 cells with phenotypic features in common with αβ T cells.

# References


Abbas, A. R., Baldwin, D., Ma, Y., Ouyang, W., Gurney, A., Martin, F., Fong, S., Campagne, M. van L., Godowski, P., Williams, P. M., Chan, A. C., & Clark, H. F. (2005). Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes & Immunity*, *6*(4), 319–331. https://doi.org/10.1038/sj.gene.6364173

Ahmadi, S. E., Soleymani, M., Shahriyary, F., Amirzargar, M. R., Ofoghi, M., Fattahi, M. D., & Safa, M. (2023). Viral vectors and extracellular vesicles: innate delivery systems utilized in CRISPR/Cas-mediated cancer therapy. *Cancer Gene Therapy*, *30*(7), 936–954. https://doi.org/10.1038/s41417-023-00597-z

Anderson, A. C., Joller, N., & Kuchroo, V. K. (2016). Lag-3, Tim-3, and TIGIT: Co-inhibitory Receptors with Specialized Functions in Immune Regulation. *Immunity*, *44*(5), 989–1004. https://doi.org/10.1016/j.immuni.2016.05.001

Arslan, A. D., Sassano, A., Saleiro, D., Lisowski, P., Kosciuczuk, E. M., Fischietti, M., Eckerdt, F., Fish, E. N., & Platanias, L. C. (2017). Human SLFN5 is a transcriptional co-repressor of STAT1-mediated interferon responses and promotes the malignant phenotype in glioblastoma. *Oncogene*, *36*(43), 6006–6019. https://doi.org/10.1038/onc.2017.205

Bangs, S. C., Baban, D., Cattan, H. J., Li, C. K.-F., McMichael, A. J., & Xu, X.-N. (2009). Human CD4+ Memory T Cells Are Preferential Targets for Bystander Activation and Apoptosis. *The Journal of Immunology*, *182*(4), 1962–1971. https://doi.org/10.4049/jimmunol.0802596

Belk, J. A., Yao, W., Ly, N., Freitas, K. A., Chen, Y.-T., Shi, Q., Valencia, A. M., Shifrut, E., Kale, N., Yost, K. E., Duffy, C. V., Daniel, B., Hwee, M. A., Miao, Z., Ashworth, A., Mackall, C. L., Marson, A., Carnevale, J., Vardhana, S. A., & Satpathy, A. T. (2022). Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors

that limit T cell persistence. *Cancer Cell*, *40*(7), 768-786.e7. https://doi.org/10.1016/j.ccell.2022.06.001

Beltra, J.-C., Manne, S., Abdel-Hakeem, M. S., Kurachi, M., Giles, J. R., Chen, Z., Casella, V., Ngiow, S. F., Khan, O., Huang, Y. J., Yan, P., Nzingha, K., Xu, W., Amaravadi, R. K., Xu, X., Karakousis, G. C., Mitchell, T. C., Schuchter, L. M., Huang, A. C., & Wherry, E. J. (2020). Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity*, *52*(5), 825-841.e8. https://doi.org/10.1016/j.immuni.2020.04.014

Bengsch, B., & Wherry, E. J. (2015). The Importance of Cooperation: Partnerless NFAT Induces T Cell Exhaustion. *Immunity*, *42*(2), 203–205. https://doi.org/10.1016/j.immuni.2015.01.023

Benyamine, A., Roy, A. L., Mamessier, E., Gertner-Dardenne, J., Castanier, C., Orlanducci, F., Pouyet, L., Goubard, A., Collette, Y., Vey, N., Scotet, E., Castellano, R., & Olive, D. (2016). BTN3A molecules considerably improve Vγ9Vδ2T cells-based immunotherapy in acute myeloid leukemia. *OncoImmunology*, *5*(10), e1146843. https://doi.org/10.1080/2162402x.2016.1146843

Bizzotto, J., Sanchis, P., Abbate, M., Lage-Vickers, S., Lavignolle, R., Toro, A., Olszevicki, S., Sabater, A., Cascardo, F., Vazquez, E., Cotignola, J., & Gueron, G. (2020). SARS-CoV-2 Infection Boosts MX1 Antiviral Effector in COVID-19 Patients. *iScience*, *23*(10), 101585. https://doi.org/10.1016/j.isci.2020.101585

Blank, C. U., Haining, W. N., Held, W., Hogan, P. G., Kallies, A., Lugli, E., Lynn, R. C., Philip, M., Rao, A., Restifo, N. P., Schietinger, A., Schumacher, T. N., Schwartzberg, P. L., Sharpe, A. H., Speiser, D. E., Wherry, E. J., Youngblood, B. A., & Zehn, D. (2019). Defining 'T cell exhaustion.' *Nature Reviews Immunology*, *19*(11), 665–674. https://doi.org/10.1038/s41577-019-0221-9

Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B., Stanley, G., Chen, S., Garnett, M., Li, W., Moffat, J., Qi, L. S.,

Shapiro, R. S., Shendure, J., Weissman, J. S., & Zhuang, X. (2022). High-content CRISPR screening. Nature Reviews Methods Primers, 2(1), 9. https://doi.org/10.1038/s43586-022-00098-7

Boucher, J. C., Yu, B., Li, G., Shrestha, B., Sallman, D., Landin, A. M., Cox, C., Karyampudi, K., Anasetti, C., Davila, M. L., & Bejanyan, N. (2023). Large Scale Ex Vivo Expansion of γδ T cells Using Artificial Antigen-presenting Cells. Journal of Immunotherapy, 46(1), 5–13. https://doi.org/10.1097/cji.0000000000000445

Brandes, M., Willimann, K., & Moser, B. (2005). Professional Antigen-Presentation Function by Human γδ T Cells. *Science*, *309*(5732), 264–268. https://doi.org/10.1126/science.1110267

Burnham, R. E., Zoine, J. T., Story, J. Y., Garimalla, S. N., Gibson, G., Rae, A., Williams, E., Bixby, L., Archer, D., Doering, C. B., & Spencer, H. T. (2020). Characterization of Donor Variability for γδ T Cell ex vivo Expansion and Development of an Allogeneic γδ T Cell Immunotherapy. *Frontiers in Medicine*, *7*, 588453. https://doi.org/10.3389/fmed.2020.588453

Cano, C. E., Pasero, C., Gassart, A. D., Kerneur, C., Gabriac, M., Fullana, M., Granarolo, E., Hoet, R., Scotet, E., Rafia, C., Hermman, T., Imbert, C., Gorvel, L., Vey, N., Briantais, A., Floch, A. C. le, & Olive, D. (2021). BTN2A1, an immune checkpoint targeting Vγ9Vδ2 T cell cytotoxicity against malignant cells. *Cell Reports*, *36*(2), 109359. https://doi.org/10.1016/j.celrep.2021.109359

Chang, S. H., & Dong, C. (2009). IL-17F: Regulation, signaling and function in inflammation. *Cytokine*, *46*(1), 7–11. https://doi.org/10.1016/j.cyto.2008.12.024

Choi, Y. J., Bowman, J. W., & Jung, J. U. (2018). A Talented Duo: IFIT1 and IFIT3 Patrol Viral RNA Caps. *Immunity*, *48*(3), 474–476. https://doi.org/10.1016/j.immuni.2018.03.001

Chuai, G., Ma, H., Yan, J., Chen, M., Hong, N., Xue, D., Zhou, C., Zhu, C., Chen, K., Duan, B., Gu, F., Qu, S., Huang, D., Wei, J., & Liu, Q. (2018). DeepCRISPR: optimizedoptimised CRISPR guide RNA design by deep learning. Genome biology, 19(1), 80. https://doi.org/10.1186/s13059-018-1459-4

Cipolletta, D., Feuerer, M., Li, A., Kamei, N., Lee, J., Shoelson, S. E., Benoist, C., & Mathis, D. (2012). PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue Treg cells. *Nature*, *486*(7404), 549–553. https://doi.org/10.1038/nature11132

Collins, S. M., Alexander, K. A., Lundh, S., Dimitri, A. J., Zhang, Z., Good, C. R., Fraietta, J. A., & Berger, S. L. (2023). TOX2 coordinates with TET2 to positively regulate central memory differentiation in human CAR T cells. *Science Advances*, *9*(29), eadh2605. https://doi.org/10.1126/sciadv.adh2605

Crawford, R. R., Prescott, E. T., Sylvester, C. F., Higdon, A. N., Shan, J., Kilberg, M. S., & Mungrue, I. N. (2015). Human CHAC1 Protein Degrades Glutathione, and mRNA Induction Is Regulated by the Transcription Factors ATF4 and ATF3 and a Bipartite ATF/CRE Regulatory Element*. *Journal of Biological Chemistry*, *290*(25), 15878–15891. https://doi.org/10.1074/jbc.m114.635144

Doering, T. A., Crawford, A., Angelosanto, J. M., Paley, M. A., Ziegler, C. G., & Wherry, E. J. (2012). Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8+ T Cell Exhaustion versus Memory. *Immunity*, *37*(6), 1130–1144. https://doi.org/10.1016/j.immuni.2012.08.021

Durant, L., Watford, W. T., Ramos, H. L., Laurence, A., Vahedi, G., Wei, L., Takahashi, H., Sun, H.-W., Kanno, Y., Powrie, F., & O'Shea, J. J. (2010). Diverse Targets of the Transcription Factor STAT3 Contribute to T Cell Pathogenicity and Homeostasis. *Immunity*, *32*(5), 605–615. https://doi.org/10.1016/j.immuni.2010.05.003

Egawa, T., & Littman, D. R. (2011). Transcription factor AP4 modulates reversible and epigenetic silencing of the Cd4 gene. *Proceedings of the National Academy of Sciences*, *108*(36), 14873–14878. https://doi.org/10.1073/pnas.1112293108

Funes, S. C., Rios, M., Fernández-Fierro, A., Covián, C., Bueno, S. M., Riedel, C. A., Mackern-Oberti, J. P., & Kalergis, A. M. (2020). Naturally Derived Heme-Oxygenase 1 Inducers and Their Therapeutic Application to Immune-Mediated Diseases. *Frontiers in Immunology*, *11*, 1467. https://doi.org/10.3389/fimmu.2020.01467

Glinos, D. A., Soskic, B., Williams, C., Kennedy, A., Jostins, L., Sansom, D. M., & Trynka, G. (2020). Genomic profiling of T-cell activation suggests increased sensitivity of memory T cells to CD28 costimulation. Genes & Immunity, 21(6–8), 390–408. https://doi.org/10.1038/s41435-020-00118-0

Grebinoski, S., Zhang, Q., Cillo, A. R., Manne, S., Xiao, H., Brunazzi, E. A., Tabib, T., Cardello, C., Lian, C. G., Murphy, G. F., Lafyatis, R., Wherry, E. J., Das, J., Workman, C. J., & Vignali, D. A. A. (2022). Autoreactive CD8+ T cells are restrained by an exhaustion-like program that is maintained by LAG3. *Nature immunology*, *23*(6), 868–877. https://doi.org/10.1038/s41590-022-01210-5

Haecker, G., & Wagner, H. (1994). Proliferative and cytolytic responses of human gamma delta T cells display a distinct specificity pattern. *Immunology*, *81*(4), 564–568.

Hale, J. S., Youngblood, B., Latner, D. R., Mohammed, A. U. R., Ye, L., Akondy, R. S., Wu, T., Iyer, S. S., & Ahmed, R. (2013). Distinct Memory CD4+ T Cells with Commitment to T Follicular Helper- and T Helper 1-Cell Lineages Are Generated after Acute Viral Infection. *Immunity*, *38*(4), 805–817. https://doi.org/10.1016/j.immuni.2013.02.020

He, B., Xing, S., Chen, C., Gao, P., Teng, L., Shan, Q., Gullicksrud, J. A., Martin, M. D., Yu, S., Harty, J. T., Badovinac, V. P., Tan, K., & Xue, H.-H. (2016). CD8+ T Cells Utilize Highly Dynamic Enhancer Repertoires and Regulatory Circuitry in Response to Infections. *Immunity*, *45*(6), 1341–1354. https://doi.org/10.1016/j.immuni.2016.11.009

Himoudi, N., Morgenstern, D. A., Yan, M., Vernay, B., Saraiva, L., Wu, Y., Cohen, C. J., Gustafsson, K., & Anderson, J. (2012). Human γδ T Lymphocytes Are Licensed for Professional Antigen Presentation by Interaction with Opsonized Target Cells. *The Journal of Immunology*, *188*(4), 1708–1716. https://doi.org/10.4049/jimmunol.1102654

Hodgson, R., Xu, X., Anzilotti, C., Deobagkar-Lele, M., Crockford, T. L., Kepple, J. D., Cawthorne, E., Bhandari, A., Cebrian-Serrano, A., Wilcock, M. J., Davies, B., Cornall, R. J., & Bull, K. R. (2022). NDRG1 is induced by antigen-receptor signaling but dispensable for B

and T cell self-tolerance. *Communications Biology*, *5*(1), 1216. https://doi.org/10.1038/s42003-022-04118-w

Hsu, T.-H., Jiang, S.-Y., Chang, W.-L., Chan, W.-L., Eckert, R. L., Scharadin, T. M., & Chang, T.-C. (2015). Involvement of RARRES3 in the regulation of Wnt proteins acylation and signaling activities in human breast cancer cells. *Cell Death & Differentiation*, *22*(5), 801–814. https://doi.org/10.1038/cdd.2014.175

Iwata, H., Goettsch, C., Sharma, A., Ricchiuto, P., Goh, W. W. B., Halu, A., Yamada, I., Yoshida, H., Hara, T., Wei, M., Inoue, N., Fukuda, D., Mojcher, A., Mattson, P. C., Barabási, A.-L., Boothby, M., Aikawa, E., Singh, S. A., & Aikawa, M. (2016). PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. *Nature Communications*, *7*(1), 12849. https://doi.org/10.1038/ncomms12849

Jang, J.-S., Lee, J.-H., Jung, N.-C., Choi, S.-Y., Park, S.-Y., Yoo, J.-Y., Song, J.-Y., Seo, H. G., Lee, H. S., & Lim, D.-S. (2018). Rsad2 is necessary for mouse dendritic cell maturation via the IRF7-mediated signaling pathway. *Cell Death & Disease*, *9*(8), 823. https://doi.org/10.1038/s41419-018-0889-y

Jeffrey, K. L., Brummer, T., Rolph, M. S., Liu, S. M., Callejas, N. A., Grumont, R. J., Gillieron, C., Mackay, F., Grey, S., Camps, M., Rommel, C., Gerondakis, S. D., & Mackay, C. R. (2006). Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nature Immunology*, *7*(3), 274–283. https://doi.org/10.1038/ni1310

Kallies, A., Zehn, D., & Utzschneider, D. T. (2020). Precursor exhausted T cells: key to successful immunotherapy? *Nature Reviews Immunology*, *20*(2), 128–136. https://doi.org/10.1038/s41577-019-0223-7

Karunakaran, M. M., Willcox, C. R., Salim, M., Paletta, D., Fichtner, A. S., Noll, A., Starick, L., Nöhren, A., Begley, C. R., Berwick, K. A., Chaleil, R. A. G., Pitard, V., Déchanet-Merville, J., Bates, P. A., Kimmel, B., Knowles, T. J., Kunzmann, V., Walter, L., Jeeves, M., … Herrmann, T. (2020). Butyrophilin-2A1 Directly Binds Germline-Encoded Regions of the

Vγ9Vδ2 TCR and Is Essential for Phosphoantigen Sensing. *Immunity*, *52*(3), 487-498.e6. https://doi.org/10.1016/j.immuni.2020.02.014

Khan, O., Giles, J. R., McDonald, S., Manne, S., Ngiow, S. F., Patel, K. P., Werner, M. T., Huang, A. C., Alexander, K. A., Wu, J. E., Attanasio, J., Yan, P., George, S. M., Bengsch, B., Staupe, R. P., Donahue, G., Xu, W., Amaravadi, R. K., Xu, X., … Wherry, E. J. (2019). TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature*, *571*(7764), 211–218. https://doi.org/10.1038/s41586-019-1325-x

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H. R., Hwang, J., Kim, J.-I., & Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nature Methods, 12(3), 237–243. https://doi.org/10.1038/nmeth.3284

Kim, D., Luk, K., Wolfe, S. A., & Kim, J.-S. (2019). Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases. Annual review of biochemistry, 88(1), 1–30. https://doi.org/10.1146/annurev-biochem-013118-111730

Kim, J., Noh, S. H., Piao, H., Kim, D. H., Kim, K., Cha, J. S., Chung, W. Y., Cho, H., Kim, J. Y., & Lee, M. G. (2016). Monomerization and ER Relocalization of GRASP Is a Requisite for Unconventional Secretion of CFTR. *Traffic*, *17*(7), 733–753. https://doi.org/10.1111/tra.12403

Kubo, S., Kataria, R., Yao, Y., Gabrielski, J. Q., Zheng, L., Markowitz, T. E., Chan, W., Song, J., Boddapati, A. K., Saeki, K., Häupl, B., Park, A. Y., Cheng, Y. H., Cui, J., Oellerich, T., & Lenardo, M. J. (2022). Early B cell factor 4 modulates FAS-mediated apoptosis and promotes cytotoxic function in human immune cells. *Proceedings of the National Academy of Sciences*, *119*(33), e2208522119. https://doi.org/10.1073/pnas.2208522119

Kumar, S., Ingle, H., Mishra, S., Mahla, R. S., Kumar, A., Kawai, T., Akira, S., Takaoka, A., Raut, A. A., & Kumar, H. (2015). IPS-1 differentially induces TRAIL, BCL2, BIRC3 and

PRKCE in type I interferons-dependent and -independent anticancer activity. *Cell Death & Disease*, *6*(5), e1758–e1758. https://doi.org/10.1038/cddis.2015.122

Kuo, A., Zhong, C., Lane, W. S., & Derynck, R. (2000). Transmembrane transforming growth factor-α tethers to the PDZ domain-containing, Golgi membrane-associated protein p59/GRASP55. *The EMBO Journal*, *19*(23), 6427–6439. https://doi.org/10.1093/emboj/19.23.6427

Lança, T., Costa, M. F., Gonçalves-Sousa, N., Rei, M., Grosso, A. R., Penido, C., & Silva-Santos, B. (2013). Protective Role of the Inflammatory CCR2/CCL2 Chemokine Pathway through Recruitment of Type 1 Cytotoxic γδ T Lymphocytes to Tumor Beds. *The Journal of Immunology*, *190*(12), 6673–6680. https://doi.org/10.4049/jimmunol.1300434

Legut, M., Gajic, Z., Guarino, M., Daniloski, Z., Rahman, J. A., Xue, X., Lu, C., Lu, L., Mimitou, E. P., Hao, S., Davoli, T., Diefenbach, C., Smibert, P., & Sanjana, N. E. (2022). A genome-scale screen for synthetic drivers of T cell proliferation. *Nature*, *603*(7902), 728–735. https://doi.org/10.1038/s41586-022-04494-7

Lei, Y., Huang, Y., Lin, J., Sun, S., Che, K., Shen, J., Liao, J., Chen, Y., Chen, K., Lin, Z., & Lin, X. (2022). Mxi1 participates in the progression of lung cancer via the microRNA-300/KLF9/GADD34 Axis. *Cell Death & Disease*, *13*(5), 425. https://doi.org/10.1038/s41419-022-04778-w

Li, J., Witten, D. M., Johnstone, I. M., & Tibshirani, R. (2012). Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics*, *13*(3), 523–538. https://doi.org/10.1093/biostatistics/kxr031

Liu, X., Wang, Y., Lu, H., Li, J., Yan, X., Xiao, M., Hao, J., Alekseev, A., Khong, H., Chen, T., Huang, R., Wu, J., Zhao, Q., Wu, Q., Xu, S., Wang, X., Jin, W., Yu, S., Wang, Y., … Dong, C. (2019). Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature*, *567*(7749), 525–529. https://doi.org/10.1038/s41586-019-0979-8

Lund, R., Aittokallio, T., Nevalainen, O., & Lahesmaa, R. (2003). Identification of Novel Genes Regulated by IL-12, IL-4, or TGF-β during the Early Polarization of CD4+ Lymphocytes. *The Journal of Immunology*, *171*(10), 5328–5336. https://doi.org/10.4049/jimmunol.171.10.5328

Lynn, R. C., Weber, E. W., Sotillo, E., Gennert, D., Xu, P., Good, Z., Anbunathan, H., Lattin, J., Jones, R., Tieu, V., Nagaraja, S., Granja, J., Bourcy, C. F. A. de, Majzner, R., Satpathy, A. T., Quake, S. R., Monje, M., Chang, H. Y., & Mackall, C. L. (2019). c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature*, *576*(7786), 293–300. https://doi.org/10.1038/s41586-019-1805-z

Mann, T. H., & Kaech, S. M. (2019). Tick-TOX, it's time for T cell exhaustion. *Nature Immunology*, *20*(9), 1092–1094. https://doi.org/10.1038/s41590-019-0478-y

Marigo, I., Zilio, S., Desantis, G., Mlecnik, B., Agnellini, A. H. R., Ugel, S., Sasso, M. S., Qualls, J. E., Kratochvill, F., Zanovello, P., Molon, B., Ries, C. H., Runza, V., Hoves, S., Bilocq, A. M., Bindea, G., Mazza, E. M. C., Bicciato, S., Galon, J., … Bronte, V. (2016). T Cell Cancer Therapy Requires CD40-CD40L Activation of Tumor Necrosis Factor and Inducible Nitric-Oxide-Synthase-Producing Dendritic Cells. *Cancer Cell*, *30*(3), 377–390. https://doi.org/10.1016/j.ccell.2016.08.004

McComb, S., Aguadé-Gorgorió, J., Harder, L., Marovca, B., Cario, G., Eckert, C., Schrappe, M., Stanulla, M., Stackelberg, A. von, Bourquin, J.-P., & Bornhauser, B. C. (2016). Activation of concurrent apoptosis and necroptosis by SMAC mimetics for the treatment of refractory and relapsed ALL. *Science Translational Medicine*, *8*(339), 339ra70. https://doi.org/10.1126/scitranslmed.aad2986

Meng, Z., Geng, X., Lin, X., Wang, Z., Chen, D., Liang, H., Zhu, Y., & Sui, Y. (2023). A prospective diagnostic and prognostic biomarker for hepatocellular carcinoma that functions in glucose metabolism regulation: Solute carrier family 37 member 3. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1869*(4), 166661. https://doi.org/10.1016/j.bbadis.2023.166661

Min, L., Isa, S. A. B. M., Fam, W. N., Sze, S. K., Beretta, O., Mortellaro, A., & Ruedl, C. (2012). Synergism between Curdlan and GM-CSF Confers a Strong Inflammatory Signature to Dendritic Cells. *The Journal of Immunology*, *188*(4), 1789–1798. https://doi.org/10.4049/jimmunol.1101755

Nishida, J., Miyazono, K., & Ehata, S. (2018). Decreased TGFBR3/betaglycan expression enhances the metastatic abilities of renal cell carcinoma cells through TGF-β-dependent and -independent mechanisms. *Oncogene*, *37*(16), 2197–2212. https://doi.org/10.1038/s41388-017-0084-0

Perkins, R. S., Singh, R., Abell, A. N., Krum, S. A., & Miranda-Carboni, G. A. (2023). The role of WNT10B in physiology and disease: A 10-year update. *Frontiers in Cell and Developmental Biology*, *11*, 1120365. https://doi.org/10.3389/fcell.2023.1120365

Piao, H., Kim, J., Noh, S. H., Kweon, H.-S., Kim, J. Y., & Lee, M. G. (2017). Sec16A is critical for both conventional and unconventional secretion of CFTR. *Scientific Reports*, *7*(1), 39887. https://doi.org/10.1038/srep39887

Piccaluga, P. P., Rossi, M., Agostinelli, C., Ricci, F., Gazzola, A., Righi, S., Fuligni, F., Laginestra, M. A., Mancini, M., Sapienza, M. R., Renzo, A. D., Tazzari, P. L., Gibellini, D., Went, P., Alviano, F., Zinzani, P. L., Bagnara, G. P., Inghirami, G., Tripodo, C., & Pileri, S. A. (2014). Platelet-derived growth factor alpha mediates the proliferation of peripheral T-cell lymphoma cells via an autocrine regulatory pathway. *Leukemia*, *28*(8), 1687–1697. https://doi.org/10.1038/leu.2014.50

Rajabzadeh, A., Hamidieh, A. A., & Rahbarizadeh, F. (2021). Spinoculation and retronectin highly enhance the gene transduction efficiency of Mucin-1-specific chimeric antigen receptor (CAR) in human primary T cells. *BMC Molecular and Cell Biology*, *22*(1), 57. https://doi.org/10.1186/s12860-021-00397-z

Roche, O., Fernández-Aroca, D. M., Arconada-Luque, E., García-Flores, N., Mellor, L. F., Ruiz-Hidalgo, M. J., & Sánchez-Prieto, R. (2020). p38β and Cancer: The Beginning of the Road.

*International Journal of Molecular Sciences*, *21*(20), 7524. https://doi.org/10.3390/ijms21207524

Seo, H., Chen, J., González-Avalos, E., Samaniego-Castruita, D., Das, A., Wang, Y. H., López-Moyado, I. F., Georges, R. O., Zhang, W., Onodera, A., Wu, C.-J., Lu, L.-F., Hogan, P. G., Bhandoola, A., & Rao, A. (2019). TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8 + T cell exhaustion. *Proceedings of the National Academy of Sciences*, *116*(25), 12410–12415. https://doi.org/10.1073/pnas.1905675116

Seo, H., González-Avalos, E., Zhang, W., Ramchandani, P., Yang, C., Lio, C.-W. J., Rao, A., & Hogan, P. G. (2021). BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nature immunology*, *22*(8), 983–995. https://doi.org/10.1038/s41590-021-00964-8

Shah, K., Al-Haidari, A., Sun, J., & Kazi, J. U. (2021). T cell receptor (TCR) signaling in health and disease. Signal Transduction and Targeted Therapy, 6(1), 412. https://doi.org/10.1038/s41392-021-00823-w

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R. A., Zhang, K., Cheng, L., & Ye, Z. (2014). Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. Cell Stem Cell, 15(1), 12–13. https://doi.org/10.1016/j.stem.2014.06.011

Stadtmauer, E. A., Fraietta, J. A., Davis, M. M., Cohen, A. D., Weber, K. L., Lancaster, E., Mangan, P. A., Kulikovskaya, I., Gupta, M., Chen, F., Tian, L., Gonzalez, V. E., Xu, J., Jung, I., Melenhorst, J. J., Plesa, G., Shea, J., Matlawski, T., Cervini, A., … June, C. H. (2020). CRISPR-engineered T cells in patients with refractory cancer. *Science*, *367*(6481). https://doi.org/10.1126/science.aba7365

Stemmer, M., Thumberger, T., Keyer, M. del S., Wittbrodt, J., & Mateo, J. L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS ONE, 10(4), e0124633. https://doi.org/10.1371/journal.pone.0124633

Tay, N. Q., Lee, D. C. P., Chua, Y. L., Prabhu, N., Gascoigne, N. R. J., & Kemeny, D. M. (2017). CD40L Expression Allows CD8+ T Cells to Promote Their Own Expansion and Differentiation through Dendritic Cells. *Frontiers in Immunology*, *8*, 1484. https://doi.org/10.3389/fimmu.2017.01484

Tijani, M., Munis, A. M., Perry, C., Sanber, K., Ferraresso, M., Mukhopadhyay, T., Themis, M., Nisoli, I., Mattiuzzo, G., Collins, M. K., & Takeuchi, Y. (2018). Lentivector Producer Cell Lines with Stably Expressed Vesiculovirus Envelopes. *Molecular Therapy - Methods & Clinical Development*, *10*, 303–312. https://doi.org/10.1016/j.omtm.2018.07.013

Tsai, S. Q., Nguyen, N. T., Malagon-Lopez, J., Topkar, V. V., Aryee, M. J., & Joung, J. K. (2017). CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. Nature Methods, 14(6), 607–614. https://doi.org/10.1038/nmeth.4278

Wang, J., Jin, J., Liang, Y., Zhang, Y., Wu, N., Fan, M., Zeng, F., & Deng, F. (2022). miR-21-5p/PRKCE axis implicated in immune infiltration and poor prognosis of kidney renal clear cell carcinoma. *Frontiers in Genetics*, *13*, 978840. https://doi.org/10.3389/fgene.2022.978840

Wherry, E. J., Ha, S.-J., Kaech, S. M., Haining, W. N., Sarkar, S., Kalia, V., Subramaniam, S., Blattman, J. N., Barber, D. L., & Ahmed, R. (2007). Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity*, *27*(4), 670–684. https://doi.org/10.1016/j.immuni.2007.09.006

Witte, M. A. de, Sarhan, D., Davis, Z., Felices, M., Vallera, D. A., Hinderlie, P., Curtsinger, J., Cooley, S., Wagner, J., Kuball, J., & Miller, J. S. (2018). Early Reconstitution of NK and γδ T Cells and Its Implication for the Design of Post-Transplant Immunotherapy. *Biology of Blood and Marrow Transplantation*, *24*(6), 1152–1162. https://doi.org/10.1016/j.bbmt.2018.02.023

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., & Zhang, B. (2014). CasOT: a genome-wide Cas9/gRNA off-target searching tool. Bioinformatics, 30(8), 1180–1182. https://doi.org/10.1093/bioinformatics/btt764

Xu, X., Chemparathy, A., Zeng, L., Kempton, H. R., Shang, S., Nakamura, M., & Qi, L. S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell*, *81*(20), 4333-4345.e4. https://doi.org/10.1016/j.molcel.2021.08.008

Yu, Z., Surface, L. E., Park, C. Y., Horlbeck, M. A., Wyant, G. A., Abu-Remaileh, M., Peterson, T. R., Sabatini, D. M., Weissman, J. S., & O'Shea, E. K. (2018). Identification of a transporter complex responsible for the cytosolic entry of nitrogen-containing bisphosphonates. *eLife*, *7*, e36620. https://doi.org/10.7554/elife.36620

Zenarruzabeitia, O., Vitallé, J., Eguizabal, C., Simhadri, V. R., & Borrego, F. (2015). The Biology and Disease Relevance of CD300a, an Inhibitory Receptor for Phosphatidylserine and Phosphatidylethanolamine. *The Journal of Immunology*, *194*(11), 5053–5060. https://doi.org/10.4049/jimmunol.1500304

Zervos, A. S., Gyuris, J., & Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell*, *72*(2), 223–232. https://doi.org/10.1016/0092-8674(93)90662-a

Zhang, Y., & Seemann, J. (2020). Rapid degradation of GRASP55 and GRASP65 reveals their immediate impact on the Golgi structure. *Journal of Cell Biology*, *220*(1), e202007052. https://doi.org/10.1083/jcb.202007052

Zhou, P., Shi, H., Huang, H., Sun, X., Yuan, S., Chapman, N. M., Connelly, J. P., Lim, S. A., Saravia, J., KC, A., Pruett-Miller, S. M., & Chi, H. (2023). Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. *Nature*, *624*(7990), 154–163. https://doi.org/10.1038/s41586-023-06733-x

# Appendix

## Full R Script for RNA Profiling analysis

```
#Start ####
#Set working directory
setwd("~/Dropbox/PhD/Bulk_RNASeq(all data)/062023_RNA seq(Genewiz)-40-876801766/")

#Load libraries
install.packages("ggrepel")
install.packages("ggplot2")
install.packages("reshape2")
install.packages("amap")
install.packages("BiocManager")
BiocManager::install("clusterProfiler")
BiocManager::install("org.Hs.eg.db") #Homo sapien organism database

library(ggplot2)
library(ggrepel)
library(reshape2)
library("amap")
library(clusterProfiler)
library(org.Hs.eg.db)


#load files for parsing ####
#Make the following files:
#Differential expression (DE)
#Expression Matrix (EM)
#Sample sheet (SS) with sample name and its group
#Load data files ####
DE.up2 <- read.csv("~/Dropbox/PhD/Bulk_RNASeq(all data)/062023_RNA seq(Genewiz)-40-876801766/DE.csv")
EM.up2 <- read.csv("~/Dropbox/PhD/Bulk_RNASeq(all data)/062023_RNA seq(Genewiz)-40-876801766/EM.csv")
SS2 <- read.csv("~/Dropbox/PhD/Bulk_RNASeq(all data)/062023_RNA seq(Genewiz)-40-876801766/SS.csv")

rm() #To remove any data from environment

#Parse data files ####
#Merge tables to have all info in 1 table
```

```
masterUP2 = merge(EM.up2, DE.up2[,2:4], by.x=0, by.y=0)
masterUP2 = masterUP2[,-1] # Get rid of first column
#reset the BOLD row IDs of master to be the Ensembl gene Ids again
#row.names (masterUP2) = masterUP2[,1]
#uniq_name <- make.names(masterUP2$ID, unique = TRUE)
#uniq_name <- make.names(rows, unique = TRUE)

#Isolate rows to identify invalid string
rows = masterUP2$ID
#set unique names for rows
uniq_name <- make.names(rows, unique = TRUE)

#Error during setting unique names
#Error: wrapup: invalid multibyte string 259

#row 259 has symbols not accepted by R
# remove row 259 and other row numbers is needed
masterUP2 = masterUP2[-259,]
masterUP2 = masterUP2[-812,]

#Set the unique Id names to data frame masterUP2 (the duplicate names will just have .1
next to it )
row.names (masterUP2) = uniq_name

class(masterUP2) = "data.frame"


#Remove first column as the names are transferred to row names now
masterUP2 = masterUP2[,-1] # Get rid of first column

#Change row IDs to gene symbols #IDs are in the 11th column #already done in file
#row.names(master) = master[,11]

#Change 1st column name  #Not applicable here
#names(master)[1] = "Ensembl_Gene_ID"

#Remove rows that have NAs from table
masterUP2 = na.omit(masterUP2)

#Sort master by padj
##Order the indexes and stores in a vector
sorted_orderUP2 = order(masterUP2["padj"], decreasing=FALSE)
##Use indexes to sort actual rows
masterUP2 = masterUP2[sorted_orderUP2,]
```

```
#Make a new column in your master table for mean expression
#Use only columns with numerical values 1:8
masterUP2$rowMeansZol = rowMeans(masterUP2[,1:4])
masterUP2$rowMeansOKT3 = rowMeans(masterUP2[,5:8])


# Add a column for -log10p to the master table (can't start with a number or symbol so we use m instead of -)
masterUP2$mlog10p = -log10(masterUP2$padj) # gets -log10p for each value

#Add a column flagging significance to the master table.
### returns a factor of TRUE / FALSE for gene significance
masterUP2$sig = as.factor(masterUP2$padj < 0.05 & abs(masterUP2$log2FoldChange) > 1.0)


#Make a scaled expression matrix
#Make EM with symbols instead of gene ID
row.names(EM.up2) = EM.up2 [,1]
EM.up2 = masterUP2[1:8]
#masterUP[ , as.vector(SS$SAMPLE)] # already using symbols
#EM.up.symbols = EM.up[,-1] # Get rid of first column
EM.up2.scaled = data.frame(t(t(scale(EM.up2))))

#Omit NAs
EM.up2.scaled  = na.omit(EM.up2.scaled)

#Make a list of significant genes using subset from master
masterUP2_sig = subset(masterUP2, sig==TRUE)
#select ONLY the significant gene NAMES and store them as a vector called: sig_genes
sigUP_genes = row.names(masterUP2_sig)
#Make expression tables of significant genes only
EM.up2.symbols.sig = EM.up2[sigUP_genes,]
EM.up2.scaled.sig = EM.up2.scaled[sigUP_genes,]

#Save your symbols expression table to disk.
write.table(masterUP2, file="masterUP2.csv", sep="\t")
write.table(masterUP2_sig, file="masterUP2_sig.csv", sep="\t")
write.table(EM.up2.symbols, file="EM.up2.symbols.csv", sep="\t")
write.table(EM.up2.symbols.sig, file="EM.up2.symbols.sig.csv", sep="\t")
write.table(EM.up2.scaled, file="EM.up2.scaled.csv", sep="\t")
write.table(EM.up2.scaled.sig, file="EM.up2.scaled.sig.csv", sep="\t")
```

```
#Create a custom theme ####

#Theme light

install.packages("ggthemes") # Install

library(ggthemes) # Load

theme_light(
  base_size = 11,
  base_family = "",
  base_line_size = base_size/22,
  base_rect_size = base_size/22
)


my_theme = theme(
  plot.title = element_text(size=10),
  axis.text.x = element_text(size=10),
  axis.text.y = element_text(size=10),
  axis.title.x = element_text(size=10),
  axis.title.y = element_text(size=10)
)


#Make plots (PCA, MA, VOLCANO, DENSITY, HEATMAP, BOXPLOTS, PATHWAY).####
#PCA ####
#prcomp accepts a numeric matrix, not a data frame as input.
## use as.matrix(sapply(table, as.numeric)) to convert your em_symbols into a matrix of
numerics.
as.matrix(sapply(EM.up2, as.numeric))
#Prcomp expects to see components (samples) by row, and dimensions (genes) by
column.
#transpose your matrix. E.g. pca = prcomp(t(numeric_matrix)).
pca = prcomp(t(EM.up2))
#extract the component data. This is a matrix (called x) held within the pca object.
pca_coordinates = data.frame(pca$x)


ggp = ggplot(pca_coordinates, aes(x=PC1, y=PC2, colour=SS2$SAMPLE.GROUP)) +
  geom_point() +
  scale_color_manual(values=c("turquoise3", "lightcoral", "black"), name="groups")+
  geom_text(aes(label=SS2$SAMPLE))+
  labs(title="PCA", x="PC1", y="PC2")+
```

```
  theme_light()

ggp

#Add % variance ####
vars = apply(pca$x, 2, var)
prop_x = round(vars["PC1"] / sum(vars),4) * 100
prop_y = round(vars["PC2"] / sum(vars),4) * 100
x_axis_label = paste("PC1 ", " (",prop_x, "%)",sep="")
y_axis_label = paste("PC2 ", " (",prop_y, "%)",sep="")

#Make an expression density plot for each sample ####
## COLOURS

# copy EM, we will then modify it
EM.up2.groups = EM.up2

# Run a loop where we take the first sample and set all its expression values to be the
name of the group.
# Then repeat for each sample
# Sounds weird, right?
# e.g every row in the column "gut_r1" will say "gut_r1". Etc.

for (index in 1:ncol(EM.up2))
{
  group = SS2[index,"SAMPLE.GROUP"]
  EM.up2.groups[,index] = group
}

EM.up2.groups = EM.up2.groups[,-8] #Removed a column that was wrong to have in the
matrix

install.packages("reshape2")
library(reshape2)

# finally melt the new table
EM.up2.groups = melt(EM.up2.groups, id.vars = NULL)

#Density plot ####

# melts em symbols
EM.up2.m = melt(EM.up2)

# adds the group column
```

```
EM.up2.m$groups = EM.up2.groups$value


# Density plot
ggp = ggplot(EM.up2.m, aes(x = log10(value+0.01), fill = groups)) +
  geom_density(alpha = 0.75) +
  facet_wrap(~variable, ncol=8) +
  theme_light() +
  theme(strip.background = element_rect(fill="transparent", linewidth=0), legend.position =
"none") +
  labs(x = "Expression (log10)", y = "Density")

ggp

#Violin plot ####
#Violin plot for 1 gene

#gene_data = em_symbols["TPX2",]
#gene_data = data.frame(t(gene_data))
#gene_data$sample_group = ss$SAMPLE_GROUP
#names(gene_data) = c("expression","sample_group")

candidate_genesUP = row.names(EM.up.scaled.sig)[1:25]
candidate_genesUP


#Violin for top 25 genes

candidate_genesUP2 = row.names(EM.up2.scaled.sig)[1:25]
candidate_genesUP2

#Top 20 genes for clearer graph
candidate_genesUP2 = row.names(EM.up2.scaled.sig)[1:20]
# or add genes manually by typing their names
#Candidate_genes =
c("TPX2","BIRC5","KIF4A","SLC7A5","KIF20A","CCND2","PODXL","ARRDC4","HAS2","OLR1
" )

gene_data = EM.up2[candidate_genesUP2,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")
```

```
ggp=ggplot(gene_data.m,aes(x=variable,y=value,fill=sample.group))+
 geom_violin(colour="black", fill="turquoise3", alpha=0.8)+
 theme_light()
ggp


#Jitter plot ####
ggp=ggplot(gene_data.m,aes(x=variable,y=value,colour=sample.group))+
 geom_jitter(width=0.1) # in jitter colouring the dots is with colour= not fill=
ggp


#To modify the current order of groups use levels
#gene_data$sample_group = factor(gene_data$sample_group,
levels=c("GROUP1??","GROUP2??","GROUP3??"))


#Boxplot ####
#For top 25 genes (scaled)
candidate_genes = row.names(EM.up2.scaled.sig)[1:20]
candidate_genes

candidate_genes = row.names(EM.up2.scaled.sig)[1:10]
candidate_genes

# other way of selecting genes by typing the names
Candidate_genes =
c("TSPAN33","PDGFRB","SNHG8","CTH","CHAC1","LRRC75A.AS1","GRK3","EEF1B2","CC
R7","CCR4","CCR4","NCS1","PRKCE","RPL5","PDGFA","ADM2","RPL9","TOX2","EBF4","ME
TRNL","HLA.DQA1","YBX3","LINC02446","CHAMP","MACROD1","ENPP2")
#plot 10 genes in 1 plot
Candidate_genes =
c("TSPAN33","PDGFRB","SNHG8","CTH","CHAC1","LRRC75A.AS1","GRK3","EEF1B2","CC
R7","CCR4")
#plot rest of genes in a different plot to make it look better
Candidate_genes =
c("NCS1","PRKCE","RPL5","PDGFA","ADM2","RPL9","TOX2","EBF4","METRNL","HLA.DQA1
","YBX3","LINC02446","CHAMP","MACROD1","ENPP2")

# Modify data to plot
gene_data = EM.up2.scaled.sig[Candidate_genes,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")
```

```
#Plot
ggp = ggplot(gene_data.m, aes(x=variable, y=value, fill=sample.group)) +
 geom_boxplot(size=0.5, outlier.size=0, alpha=0.5, colour="black") +
 theme_light()

ggp



#For top 25 genes (scaled)
candidate_genes = row.names(EM.up2.symbols.sig)[1:25]
candidate_genes

Candidate_genes =
c("TSPAN33","PDGFRB","SNHG8","CTH","CHAC1","LRRC75A.AS1","GRK3","EEF1B2","CC
R7","CCR4","CCR4","NCS1","PRKCE","RPL5","PDGFA","ADM2","RPL9","TOX2","EBF4","ME
TRNL","HLA.DQA1","YBX3","LINC02446","CHAMP","MACROD1","ENPP2")

#plot 10 genes in 1 plot
Candidate_genes =
c("TSPAN33","PDGFRB","SNHG8","CTH","CHAC1","LRRC75A.AS1","GRK3","EEF1B2","CC
R7","CCR4")

gene_data = EM.up2.symbols.sig[candidate_genes,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")

ggp = ggplot(gene_data.m, aes(x=variable, y=value, fill=sample.group)) +
 geom_boxplot(size=0.5, outlier.size=0, alpha=0.5, colour="black") +
 theme_light()
ggp

#plot rest of genes in a different plot to make it look better
Candidate_genes =
c("NCS1","PRKCE","RPL5","PDGFA","ADM2","RPL9","TOX2","EBF4","METRNL","HLA.DQA1
","YBX3","LINC02446","CHAMP","MACROD1","ENPP2")

gene_data = EM.up2.symbols.sig[candidate_genes,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")
```

```
ggp = ggplot(gene_data.m, aes(x=variable, y=value, fill=sample.group)) +
 geom_boxplot(size=0.5, outlier.size=0, alpha=0.5, colour="black") +
 theme_light()
ggp

##Faceted boxplot ####
#For all genes in one boxplot x=variable >> meaning gene name instead of sample_group
#plot 10 genes in 1 plot
Candidate_genes =
c("TSPAN33","PDGFRB","SNHG8","CTH","CHAC1","LRRC75A.AS1","GRK3","EEF1B2","CC
R7","CCR4")

gene_data = EM.up2.symbols.sig[Candidate_genes,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")

ggp=ggplot(gene_data.m,aes(x=sample.group,y=value,fill=sample.group))+
 geom_boxplot(size=0.5,outlier.size=0,alpha=0.5,colour="black")+
 facet_wrap(~variable, ncol=25)+#Chane ncol dpending on how many column you want in
you plots layout (10 next to each other>> ncol=10)
 theme_light()+
 theme(axis.text.x = element_text(angle = 45, hjust = 1)) # makes the x-axis at an angle 45
#must be placed after any other theme

ggp

#plot rest of genes in a different plot to make it look better
Candidate_genes =
c("NCS1","PRKCE","RPL5","PDGFA","ADM2","RPL9","TOX2","EBF4","METRNL","HLA.DQA1
","YBX3","LINC02446","CHAMP","MACROD1","ENPP2")

gene_data = EM.up2.symbols.sig[Candidate_genes,]
gene_data = data.frame(t(gene_data))
gene_data$sample.group = SS2$SAMPLE.GROUP
gene_data.m = melt(gene_data, id.vars="sample.group")

ggp=ggplot(gene_data.m,aes(x=sample.group,y=value,fill=sample.group))+
 geom_boxplot(size=0.5,outlier.size=0,alpha=0.5,colour="black")+
 facet_wrap(~variable, ncol=25)+#Chane ncol dpending on how many column you want in
you plots layout (10 next to each other>> ncol=10)
 theme_light()+
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # makes the x-axis at an angle 45
#must be placed after any other theme

ggp

#Heatmap ####

candidate_genes = row.names(EM.up2.symbols.sig)[1:50]
candidate_genes
#Steps to do for clustering
gene_data = EM.up2.scaled.sig[candidate_genes,]
hm.matrix = as.matrix(gene_data)
y.dist = Dist(hm.matrix, method="spearman")
y.cluster = hclust(y.dist, method="average")
y.dd = as.dendrogram(y.cluster)
y.dd.reorder=reorder(y.dd,0,FUN="average")
y.order = order.dendrogram(y.dd.reorder)
hm.matrix_clustered = hm.matrix[y.order,]
hm.matrix_clustered = melt(hm.matrix_clustered)

## Fix the colours
colours = c("turquoise3","lightcoral")


colorRampPalette(colours)(5000) # this means it will find 100 colours between blue and
red to make the scale

#Run ggplot for heatmap
ggp = ggplot(hm.matrix_clustered, aes(x=Var2, y=Var1, fill=value)) + geom_tile() +
  scale_fill_gradientn(colours=colorRampPalette(colours)(200))

ggp

#load all genes file for ORA as shown previously ####

#ORA analysis

sigALL2_genes = row.names(masterALL2_sig)

sigALL2_genes_names = sigALL2_genes #Make only 1 column with gene symbols (vector)
sig genes is a vector
sigALL2_genes_names
sigALL2_genes_entrez = bitr(sigALL2_genes_names, fromType = "SYMBOL", toType =
"ENTREZID", OrgDb = org.Hs.eg.db)
```

```
pathway_results=enrichGO(gene=sigALL2_genes_entrez$ENTREZID,
                OrgDb=org.Hs.eg.db, readable=T,ont="BP", pvalueCutoff=0.05,
qvalueCutoff=0.10)

ggp = barplot(pathway_results, showCategory=10)
ggp

ggp = dotplot(pathway_results, showCategory=10)
ggp

ggp = goplot(pathway_results, showCategory=10)
ggp

ggp = cnetplot(pathway_results, categorySize="pvalue")
ggp

#
pathway_results=enrichGO(gene=sigALL2_genes_entrez$ENTREZID,
                OrgDb=org.Hs.eg.db, readable=T,ont="MF", pvalueCutoff=0.05,
qvalueCutoff=0.10)

ggp = barplot(pathway_results, showCategory=10)
ggp

ggp = dotplot(pathway_results, showCategory=10)
ggp

ggp = goplot(pathway_results, showCategory=10)
ggp

ggp = cnetplot(pathway_results, categorySize="pvalue")
ggp



#Plot GSEA ####


if (!("clusterProfiler" %in% installed.packages())) {
  # Install this package if it isn't installed yet
  BiocManager::install("clusterProfiler", update = FALSE)
}

if (!("msigdbr" %in% installed.packages())) {
```

```
  # Install this package if it isn't installed yet
  BiocManager::install("msigdbr", update = FALSE)
}

if (!("org.Hs.eg.db" %in% installed.packages())) {
  # Install this package if it isn't installed yet
  BiocManager::install("org.Hs.eg.db", update = FALSE)
}

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("msigdb")

a
Yes
library(msigdbr)
library(ExperimentHub)
library(GSEABase)

if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("ExperimentHub")

ExperimentHub()
a
Yes

install.packages("dbplyr")
library(dbplyr)

#Error in finding function ExperimentHub()

#A solution
readLines("https://experimenthub.bioconductor.org")

eh = ExperimentHub()
query(eh , 'msigdb')

library(fgsea)

#msigdbr_df <- msigdbr(species = "human", category = "H")
```

```
#head(msigdbr_df)

hs_hallmark_sets <- msigdbr(
  species = "Homo sapiens", # Replace with species name relevant to your data
  category = "H" )


# fixing format to work with fgsea
#pathwaysH = split(x = msigdbr_df$entrez_gene, f = msigdbr_df$gs_name)
pathwaysH = split(x = hs_hallmark_sets$entrez_gene, f = hs_hallmark_sets$gs_name)


# run fgsea enrichment
#fgseaRes <- fgsea(pathways=pathwaysH, ranks, ..)

fgseaRes <- fgsea(pathways=pathwaysH,stats = exampleRanks,
              minSize=15,
              maxSize=500)

# Let's create a named vector ranked based on the log2 fold change values
lfc_vector <- masterALL2$log2FoldChange
names(lfc_vector) <- rownames(masterALL2)

# We need to sort the log2 fold change values in descending order here
lfc_vector <- sort(lfc_vector, decreasing = TRUE)

head(lfc_vector)

set.seed(2020) # Set the seed so our results are reproducible

library(clusterProfiler)


gsea_results_h <- GSEA(
  geneList = lfc_vector, # Ordered ranked gene list
  minGSSize = 25, # Minimum gene set size
  maxGSSize = 500, # Maximum gene set set
  pvalueCutoff = 0.05, # p-value cutoff
  eps = 0, # Boundary for calculating the p value
  seed = TRUE, # Set seed to make results reproducible
  pAdjustMethod = "BH", # Benjamini-Hochberg correction
  TERM2GENE = dplyr::select(
    hs_hallmark_sets,
    gs_name,
```

```
  gene_symbol
 )
)

gsea_result_df_h <- data.frame(gsea_results_h@result)

# Install and load the gt package to make a table
install.packages("gt")
library(gt)

# Display the table >> use the gsea data frame format for tables
#gsea_result_df <- data.frame(gsea_results@result)

gt(gsea_result_df)

library(ggplot2)
library(gridExtra)

# Create a table graphic
table_grob <- tableGrob(results_df)

# Save as high-resolution PNG
ggsave(filename = "gsea_results.png", plot = table_grob, width = 20, height = 15, dpi = 300)

#If table is too large to view all >>> Remove last column
results_df <- gsea_result_df[, -ncol(gsea_result_df)]
#Still too large to view >>> Remove the last column again to make it even smaller
results_df <- results_df[, -ncol(results_df)]




#Save results table
write.table(gsea_result_df_h, file="gsea_result_df_h.csv", sep="\t")

#If the tabs are not seperated after extracting (the file has commas) >> use this format to
save as tab-separated values (TSV) file
write.table(gsea_result_df_h, "gsea_result_df_h.txt", sep = "\t", row.names = FALSE, quote
= FALSE)


##for immune-signatures data set C7 from msigDB
#Install required packages
```

```
install.packages("msigdbr")
library(msigdbr)

#Get the gene set
c7_df <- msigdbr(species = "Homo sapiens", category = "C7")

#The resulting c7_df dataframe will have columns for the gene set name (gs_name), gene
(gene_symbol), and other metadata.
#You can then use this dataframe as the TERM2GENE input in your GSEA function:

TERM2GENE = dplyr::select(c7_df, gs_name, gene_symbol)

gsea_results_C7 <- GSEA(
  geneList = lfc_vector,
  minGSSize = 25,
  maxGSSize = 500,
  pvalueCutoff = 0.05,
  eps = 0,
  seed = TRUE,
  pAdjustMethod = "BH",
  TERM2GENE = TERM2GENE
)

gsea_result_C7_df <- data.frame(gsea_results_C7@result)

#Save a csv table
write.table(gsea_result_C7_df, file="gsea_result_C7_df.csv", sep="\t")

#If the tabs are not seperated after extracting (the file has commas) >> use this format to
save as tab-separated values (TSV) file
write.table(gsea_result_C7_df, "gsea_result_C7_df.txt", sep = "\t", row.names = FALSE,
quote = FALSE)

# Install and load the gt package to make a table
install.packages("gt")
library(gt)

# Display the table >> use the gsea data frame format for tables
#gsea_result_df <- data.frame(gsea_results@result)

gt(gsea_result_C7_df)

library(ggplot2)
```

```
library(gridExtra)

# Create a table graphic
table_grob <- tableGrob(results_df_C7)

# Save as high-resolution PNG
ggsave(filename = "gsea_results_C7.png", plot = table_grob, width = 50, height = 50, dpi =
300, limitsize = FALSE)

#If table is too large to view all >>> Remove last column
results_df_C7 <- gsea_result_C7_df[, -ncol(gsea_result_C7_df)]
#Still too large to view >>> Remove the last column again to make it even smaller
results_df <- results_df[, -ncol(results_df)]
#save again


#plot enrichment maps interactive html file ####
# Install and load the enrichplot package
install.packages("enrichplot")
library(enrichplot)

# Assuming you have an enrichment result object named 'gsea_result'
# This can be a result from clusterProfiler or other supported formats
emapplot(gsea_results)

head(gsea_results)

#Visualize C7 interactive network
install.packages("networkD3")
library(networkD3)

# Sample Data
data <- data.frame(gsea_results_C7)

# Create a node list
nodes <- data.frame(name = data$ID)

# Create an edge list
edges <- data.frame(source = rep(0:(nrow(data)-1), each = nrow(data)),
                    target = rep(0:(nrow(data)-1), times = nrow(data)))

# Remove self-connections
edges <- subset(edges, source != target)
```

```
# Calculate edge weights based on shared genes
edges$weight <- mapply(function(s, t) {
  length(intersect(unlist(strsplit(data$core_enrichment[s+1], "/")),
                   unlist(strsplit(data$core_enrichment[t+1], "/"))))
}, s = edges$source, t = edges$target)

# Filter out edges with zero weight
edges <- subset(edges, weight > 0)


# Add a group column to the nodes data frame
nodes$group <- 1

#To get names on nodes
# Add a size column to the nodes data frame
nodes$size <- 20  # This sets a constant size for all nodes; adjust as necessary

# Generate the force directed network
network_C7 <- forceNetwork(Links = edges, Nodes = nodes,
                           Source = 'source', Target = 'target',
                           Value = 'weight', NodeID = 'name',
                           Nodesize = 'size',  # Use the size column for node sizes
                           Group = 'group', opacity = 0.9)

print(network_C7)

library(htmlwidgets)

# Save the visualization to an HTML file
saveWidget(network_C7, file = "network_C7.html")
saveWidget(network_h, file = "network_h.html")

webshot(network_h, file = "network_h.png")


wkhtmltoimage network_h.html network_h.png

wkhtmltoimage

# Enrichment map for gsea from gsea results table or data frames ####

# Plot enrichment
install.packages("BiocManager")
```

```
BiocManager::install("clusterProfiler")


a
Yes
BiocManager::install("enrichplot")
BiocManager::install("DOSE", force = TRUE)

library(clusterProfiler)
library(enrichplot)
library(DOSE) # if the tap doesn't work to complete words


gsea_result_df_h$gene <- lapply(gsea_result_df_h$core_enrichment, function(x) strsplit(x,
"/")[[1]])

#
# Calculate Jaccard similarity
jaccard_similarity <- function(list1, list2) {
  intersect_len <- length(intersect(list1, list2))
  union_len <- length(union(list1, list2))
  return(intersect_len / union_len)
}

# Compute pairwise similarities for your gene sets

similarity_matrix_h <- matrix(0, nrow=length(gsea_result_df_h$gene),
ncol=length(gsea_result_df_h$gene))
rownames(similarity_matrix_h) <- gsea_result_df_h$ID
colnames(similarity_matrix_h) <- gsea_result_df_h$ID

for (i in 1:length(gsea_result_df_h$gene)) {
  for (j in 1:length(gsea_result_df_h$gene)) {
    similarity_matrix_h[i,j] <- jaccard_similarity(gsea_result_df_h$gene[[i]],
gsea_result_C7_df_top20$gene[[j]])
  }
}


# Use the similarity matrix to plot in igraph
install.packages("igraph")

library(igraph)
```

```
g_h <- graph_from_adjacency_matrix(similarity_matrix_h, mode = "undirected", weighted =
TRUE, diag = FALSE)

plot(g_h, vertex.color="skyblue", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")


#PDF is better for reports and publications >> highest quality
pdf("enrichment_map_h.pdf", width=8, height=8)  # width and height are in inches here
plot(g_h, vertex.color="skyblue", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")

dev.off()

png("enrichment_map_h.png", width=800, height=800)
plot(g_h, vertex.color="skyblue", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")
dev.off()

#Repeat for C7 immune-signatures data sets

#convert gene names from seperated by / in one cell to a list
gsea_result_C7_df$gene <- lapply(gsea_result_C7_df$core_enrichment, function(x)
strsplit(x, "/")[[1]])


#Calculate simsilary matrix before plotting enrichment map

# Calculate Jaccard similarity
jaccard_similarity <- function(list1, list2) {
  intersect_len <- length(intersect(list1, list2))
  union_len <- length(union(list1, list2))
  return(intersect_len / union_len)
}

# Compute pairwise similarities for your gene sets (replace names for the desired gene set
likse from similarity_matrix_c7 to similarity_matrix_h and all gsea_result_C7_df)
similarity_matrix_c7 <- matrix(0, nrow=length(gsea_result_C7_df$gene),
ncol=length(gsea_result_C7_df$gene))
rownames(similarity_matrix_c7) <- gsea_result_C7_df$ID
colnames(similarity_matrix_c7) <- gsea_result_C7_df$ID

for (i in 1:length(gsea_result_C7_df$gene)) {
  for (j in 1:length(gsea_result_C7_df$gene)) {
```

```
   similarity_matrix_c7[i,j] <- jaccard_similarity(gsea_result_C7_df$gene[[i]],
gsea_result_C7_df$gene[[j]])
  }
}


#Top 20 of C7
#Note: C7 is a large list so had to do top20 only
#convert gene names from seperated by / in one cell to a list
gsea_result_C7_df$gene <- lapply(gsea_result_C7_df$core_enrichment, function(x)
strsplit(x, "/")[[1]])
gsea_result_C7_df_top20 <- gsea_result_C7_df[1:20, ]
similarity_matrix_c7_top20 <- matrix(0, nrow=length(gsea_result_C7_df_top20$gene),
ncol=length(gsea_result_C7_df_top20$gene))
rownames(similarity_matrix_c7_top20) <- gsea_result_C7_df_top20$ID
colnames(similarity_matrix_c7_top20) <- gsea_result_C7_df_top20$ID

for (i in 1:length(gsea_result_C7_df_top20$gene)) {
  for (j in 1:length(gsea_result_C7_df_top20$gene)) {
    similarity_matrix_c7_top20[i,j] <- jaccard_similarity(gsea_result_C7_df_top20$gene[[i]],
gsea_result_C7_df_top20$gene[[j]])
  }
}
#Plot using igraph (loaded above)
g_c7_top20 <- graph_from_adjacency_matrix(similarity_matrix_c7_top20, mode =
"undirected", weighted = TRUE, diag = FALSE)

plot(g_c7_top20, vertex.color="pink", vertex.size=10, vertex.label.cex = 0.2, edge.width =
E(g_c7_top20)$weight*4, edge.color = "Black")
#plot(g_c7, layout=layout_with_kk, vertex.color="pink", vertex.label.cex=0.5,
vertex.size=10, edge.width=2, edge.color = "Black")

pdf("enrichment_map_c7_top20.pdf", width=8, height=8)  # width and height are in inches
here
plot(g_c7_top20, vertex.color="pink", vertex.size=10, vertex.label.cex = 0.2, edge.width =
E(g_c7_top20)$weight*4, edge.color = "Black")
dev.off()

png("enrichment_map_c7_top20.png", width=800, height=800)
plot(g_c7_top20, vertex.color="pink", vertex.size=10, vertex.label.cex = 0.2, edge.width =
E(g_c7_top20)$weight*4, edge.color = "Black")
dev.off()
```

```
#Save all plots for all ####

#PDF is better for reports and publications >> highest quality
pdf("enrichment_map_h.pdf", width=8, height=8)  # width and height are in inches here
plot(g_h, vertex.color="skyblue", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")

dev.off()

png("enrichment_map_h.png", width=800, height=800)
plot(g_h, vertex.color="skyblue", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")
dev.off()

pdf("enrichment_map_c7.pdf", width=8, height=8)  # width and height are in inches here
plot(g_c7, vertex.color="pink", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")

dev.off()

png("enrichment_map_c7.png", width=800, height=800)
plot(g_c7, vertex.color="pink", vertex.label.cex = 0.3, edge.width = E(g_h)$weight*6,
edge.color = "Black")
dev.off()
```

## gRNA library for CRISPR knock-out screen

| gRNA library for CRISPR knock-out screen | |
|---|---|
| **sgRNA ID** | **sgRNA Sequence** |

| | |
|---|---|
| **sgA4GALT_1** | GGGGTGTCAAGGTGGGGCAG |
| **sgA4GALT_2** | CTGATGAAAGGGCTTCCGGG |
| **sgA4GALT_3** | GCCGCGTACCAGTCGGCCAG |
| **sgA4GALT_4** | GTGTGGAACAAGAAGAGCCA |
| **sgA4GALT_5** | GCAGCCGCAGCAGGAGGTCG |
| **sgABCE1_1** | GGAAGACTTAAACCTGATGA |
| **sgABCE1_2** | CAGCAGGTTTGCCCAAGCAA |
| **sgABCE1_3** | GACAATTGATAAGGCGCCAA |
| **sgABCE1_4** | AAAATCTCCTGCCAGTCAGG |
| **sgABCE1_5** | CCAGATTCCTAAGGCTGCAA |
| **sgABCF2_1** | GAAGACGCTACAGAAAATGA |
| **sgABCF2_2** | ACTTCCATCACACAATGCAA |
| **sgABCF2_3** | GAGCGGCTGGCTCATGAGGA |
| **sgABCF2_4** | GGAGCTCTACGAGCGCCTGG |
| **sgABCF2_5** | GAAGACGCGGCTAGAGCTGG |
| **sgACSL1_1** | GAAAATACCGGAACAGCTCA |
| **sgACSL1_2** | GTTTCCGAGAGCCTAAACAA |
| **sgACSL1_3** | CAAACACCACGCTGAAGCGA |
| **sgACSL1_4** | AAAGAAGCAGAGCTTCGCAG |
| **sgACSL1_5** | CAGCACAGTGGCAGACACCG |
| **sgACVR1B_1** | GAACAGGAGGAACACCGGGC |
| **sgACVR1B_2** | GTAGTCAGTGTAGCAGCAGT |
| **sgACVR1B_3** | GGAGAGACACATCTCACATG |
| **sgACVR1B_4** | GGTACCTGAGCCAGACCCTG |
| **sgACVR1B_5** | TGGTTCGGGCCACTGTGCGC |
| **sgADAM19_1** | GTCACGCTCAGCACTTGCCG |
| **sgADAM19_2** | GAGCCCCTCCCTGACAGCAA |
| **sgADAM19_3** | GGAGACACCTTTGGAAACTG |
| **sgADAM19_4** | GGTGTCAATGGGCACCGCGT |
| **sgADAM19_5** | GGGAGGCAGATCCAGTGCCG |
| **sgADO_1** | GACAAGCTAGACGCGGGCGG |
| **sgADO_2** | GCTGCAGTGTGGCCTTGCGC |
| **sgADO_3** | GCATGCCGTGCATGCCCGGG |
| **sgADO_4** | TTCGCGGGCCGAGTACACCG |
| **sgADO_5** | GGTGCGGTGTGAGGATGCAG |
| **sgAEN_1** | GCATTTGGGATGGTGAGGGA |
| **sgAEN_2** | GAAGGAGCCGACAGCACCAG |
| **sgAEN_3** | GAAAGGGGTGGGCAGTGGGG |
| **sgAEN_4** | TGTCGCTGCCCCGAAAGGGG |

| | |
|---|---|
| **sgAEN_5** | GCAGTGCCCCATGCAGCAGA |
| **sgAHCYL2_1** | GAGGATAGTACTGCTGGCAG |
| **sgAHCYL2_2** | CTGATGATGAGACATCGCCC |
| **sgAHCYL2_3** | GGAGGGCTGGCAGCCAAACA |
| **sgAHCYL2_4** | GCAGATCTTGGATGATGGAG |
| **sgAHCYL2_5** | GAAAATCAAGGGCATAGTAG |
| **sgALAS1_1** | GCTGCAGTGGACAATGCCCG |
| **sgALAS1_2** | ATAACTGCCCCACACACCCG |
| **sgALAS1_3** | AGAGTGCGGCATCTTTCCCA |
| **sgALAS1_4** | GGTGGCTGACATCATTGTGG |
| **sgALAS1_5** | GGGCTTTATGGGGCTCGAGG |
| **sgANAPC1_1** | GATGAGACTCAACATTCTGG |
| **sgANAPC1_2** | GTTGCCTGCATACATCGCAA |
| **sgANAPC1_3** | GACTTGTCTCATGAACATGA |
| **sgANAPC1_4** | CTGGAAAGGTGAAGTCACAG |
| **sgANAPC1_5** | TGTCCATAACAGAGTCACCC |
| **sgANGPTL4_1** | CAAGCACCTAGACCATGAGG |
| **sgANGPTL4_2** | GTGCTACTGAGCGCTCAGGG |
| **sgANGPTL4_3** | GGCTCTCAGGGGCTAACGGG |
| **sgANGPTL4_4** | ACAAGGTGGCCCAGCAGCAG |
| **sgANGPTL4_5** | GCAGACACTCACGGTGCAGG |
| **sgANKRD37_1** | GTTGCTGGATTGCAACCCCG |
| **sgANKRD37_2** | GGAAGTGTAGAAAATACCAG |
| **sgANKRD37_3** | GAAGCATTTGCTGGAGACAG |
| **sgANKRD37_4** | GTCCACTTAGCCGCAGGAAG |
| **sgANKRD37_5** | GCCAGAGAAGAAAGCAAGCA |
| **sgANKRD52_1** | GCAGTGGAGCTGACTTGAGG |
| **sgANKRD52_2** | GAAGCTCACTCACCCCGCGG |
| **sgANKRD52_3** | AAGATGGCCTGGACCAGGGG |
| **sgANKRD52_4** | TGAATGCCCGGGACAAGCTG |
| **sgANKRD52_5** | AGAAAAGCTGCCCAATGCAG |
| **sgAPBB2_1** | GGGATAACTCACCATGCAGG |
| **sgAPBB2_2** | ACGGAAGAATGCCAAAGCGC |
| **sgAPBB2_3** | GCTGTCAGTGAACATGAACG |
| **sgAPBB2_4** | AAACGTGGCACTCAAAGCGC |
| **sgAPBB2_5** | TATGACCTCATCCAACAAGG |
| **sgAREG_1** | AGCGACAGCACCACCGGCGC |
| **sgAREG_2** | GAAGCGTGAACCATTTTCTG |
| **sgAREG_3** | AGAAGGCATTTCACTCACAG |

| | |
|---|---|
| **sgAREG_4** | GAATATTTCGGTGAACGGTG |
| **sgAREG_5** | GAGGACGGTTCACTACTAGA |
| **sgARSB_1** | GCAGCAGCGGGAGGACGACG |
| **sgARSB_2** | GAATGCCTTCCAACCCGCCG |
| **sgARSB_3** | AGTAGGTATCAAATCCTCGG |
| **sgARSB_4** | ACGGAGGGCAGACTTTGGCA |
| **sgARSB_5** | CACACGTCGAAGCCATCCAG |
| **sgARVCF_1** | GAGGTTCTGCAGAGCGCCGG |
| **sgARVCF_2** | GCAGTCTGAGACCGACAAGG |
| **sgARVCF_3** | GGCCTCCTGCTCCTTCACCG |
| **sgARVCF_4** | GCTGGAGGAGACCGTGACGG |
| **sgARVCF_5** | GCTGCGTGGTGGTGGCCCAG |
| **sgATP2A2_1** | TTGCCGGCACAGCTTCACGG |
| **sgATP2A2_2** | GGAGAACGCGCACACCAAGA |
| **sgATP2A2_3** | AGGACAGAAAGAGTGTGCAG |
| **sgATP2A2_4** | CAAGATCCGGGATGAAATGG |
| **sgATP2A2_5** | CAAGGGTTTCCACAGACGGG |
| **sgATP9A_1** | GGTGCGCCAGAAGAAGCGGA |
| **sgATP9A_2** | GAGTGGCTGAGATGCTGCGG |
| **sgATP9A_3** | GAAGGGACTCGAAAATCCCC |
| **sgATP9A_4** | GGAGGCAAAGTAAAACACGG |
| **sgATP9A_5** | GGAGATCCGATGCTACGTGC |
| **sgATR_1** | GAAATCAAGCAACATCACGG |
| **sgATR_2** | AAGAGCACCAAAACAGACTG |
| **sgATR_3** | GCCCAGGTCACCAATTGTGG |
| **sgATR_4** | GCTACTGCCAGAGTATCCTG |
| **sgATR_5** | GAAAATGAAACCCCACCTGA |
| **sgBACH2_1** | GTGTAGGCAAACTGTAACAG |
| **sgBACH2_2** | CAGCAGAGAAAACATCCGCG |
| **sgBACH2_3** | GGATGAAGAGGAGGAGACGA |
| **sgBACH2_4** | GGAGACGATGGATTCAGAGA |
| **sgBACH2_5** | GAATGTGCTTGCAAAACCTG |
| **sgBASP1_1** | GGAGAAAGACAAGAAGGCCG |
| **sgBASP1_2** | GAAGCCCGACCAGGACGCCG |
| **sgBASP1_3** | GGAGGGCGAGAAGGACGCGG |
| **sgBASP1_4** | GAAGGAGAGTGAGCCCCAGG |
| **sgBASP1_5** | GGCGGAGCCCGAGAAGACGG |
| **sgBATF_1** | GGACTCTACCTGTTTGCCAG |
| **sgBATF_2** | CCGCCCAGAAGAGCCGACAG |

| | |
|---|---|
| **sgBATF_3** | GCTGTCGGAGCTGTGAGGCA |
| **sgBATF_4** | ATGATGTGAGAAGAGTTCAG |
| **sgBATF_5** | GAAGCTGGAGTCACTGCTGT |
| **sgBATF3_1** | GAGCGCCCGGCATGTCGCAA |
| **sgBATF3_2** | ATGATGACAGGAAGGTCCGA |
| **sgBATF3_3** | GAAGGCTGACAAGCTCCATG |
| **sgBATF3_4** | CCTCTGCAGGACGCTGCCGG |
| **sgBATF3_5** | TCGGAAGAAGCAGACCCAGA |
| **sgBCL2L1_1** | GCACCTGGCAGACAGCCCCG |
| **sgBCL2L1_2** | TGTGGAAGAGAACAGGACTG |
| **sgBCL2L1_3** | CCAGTGGCTCCATTCACCGC |
| **sgBCL2L1_4** | CAGCAGTTTGGATGCCCGGG |
| **sgBCL2L1_5** | CAGGCGACGAGTTTGAACTG |
| **sgBLVRB_1** | GGTGAGCCCGGTCTGGCCAG |
| **sgBLVRB_2** | GAACATCTCCCACTACCACG |
| **sgBLVRB_3** | GATGTGGACAAGACCGTGGC |
| **sgBLVRB_4** | TCATGGTGTGGACAAGGTCG |
| **sgBLVRB_5** | TTGCACCGCCTGCGCCAGGG |
| **sgBNIP3L_1** | AGAAGGAGAGAAGGAAGTCG |
| **sgBNIP3L_2** | TCAGGACAGAGTAGTTCCAG |
| **sgBNIP3L_3** | AAAGTGGAGCCATGAAGAAA |
| **sgBNIP3L_4** | TCACTGTTGAGGCCGGCCGG |
| **sgBNIP3L_5** | TTGTTGTGCAGGGGCGGCGG |
| **sgC17orf51_1** | AAAGTGCAGCCTGTGCACGG |
| **sgC17orf51_2** | ATGGGTGCTCCAGAAAGCTG |
| **sgC17orf51_3** | CGAGAGGCCCACACAGGACA |
| **sgC17orf51_4** | GTCCCGCAGGCGCAGTGCTG |
| **sgC17orf51_5** | GAATGGAAAAGCGATCCAGA |
| **sgC17orf89_1** | CACGCACCTGCCGTACGCCG |
| **sgC17orf89_2** | GCGGAGCTGCTTCGCCGCTG |
| **sgC17orf89_3** | GCTCATGTCGGCTAACGGAG |
| **sgC17orf89_4** | CCTTACTCAGGCGGCCGCCC |
| **sgC17orf89_5** | GCGCGCAGAAGTCCTTACTC |
| **sgCAD_1** | GAAGGAGATTGAGTACGAGG |
| **sgCAD_2** | GCAGGAGGACACCTATGTGG |
| **sgCAD_3** | GAAGCCCTGGGATGCCACGG |
| **sgCAD_4** | GCAGCCACTCATGCAGGGTG |
| **sgCAD_5** | GGGAGCTGACCAAGAAGTTG |
| **sgCADPS_1** | GGAGCGGCTGCAGAAAGAGG |

| | |
|---|---|
| **sgCADPS_2** | GTGGTGGTCAGATTTAATGG |
| **sgCADPS_3** | GGAGGAGGAGAGCGGCAAGG |
| **sgCADPS_4** | GGTGGTGAGCGAGAAGGAGA |
| **sgCADPS_5** | ACCGTGGAGAAGAGGACCCG |
| **sgCADPS2_1** | TGGACGATGAGCAGGAGCGG |
| **sgCADPS2_2** | GCAGCTTGAAGGCTATACTG |
| **sgCADPS2_3** | CAGAAACTGAATCCTAAAGG |
| **sgCADPS2_4** | GAATGAAGAGCATCATGCAG |
| **sgCADPS2_5** | GACGCTCATCAAGATTGTGA |
| **sgCALCOCO1_1** | AAGTTGACTCCACCACGGGA |
| **sgCALCOCO1_2** | TGAACTGGTGACCCTGGAGG |
| **sgCALCOCO1_3** | GATGCAGCTGAAGCTACAGC |
| **sgCALCOCO1_4** | GCAGGAGCACACGGAGCTGA |
| **sgCALCOCO1_5** | TGGGGAGATCACAGAAGAGA |
| **sgCALU_1** | TGATGGTTAGAGATGAGCGG |
| **sgCALU_2** | GAAGAGAGCAAGGAAAGGCT |
| **sgCALU_3** | GATGATCCAGATCCTGATGA |
| **sgCALU_4** | GAGCAAACCCACAGAAAAGA |
| **sgCALU_5** | TGAACCTTGTCACTGAGCTG |
| **sgCASP3_1** | GGCGTGTCATAAAATACCAG |
| **sgCASP3_2** | GTTTGTGTGCTTCTGAGCCA |
| **sgCASP3_3** | TCTAACACCCAGGCCTGCCG |
| **sgCASP3_4** | AACGTTCTCACCAGGTGCTG |
| **sgCASP3_5** | TGTGGCATTGAGACAGACAG |
| **sgCCDC6_1** | CGAGAGCGACACGGACGGGG |
| **sgCCDC6_2** | CAAGGTGCTGAAGATAGAGC |
| **sgCCDC6_3** | GAAGTGCAAGGCACTGCAGG |
| **sgCCDC6_4** | GAAAATGGCACAGTATCTGG |
| **sgCCDC6_5** | GGCCTGCTTGAACTCGGAGA |
| **sgCCL17_1** | ACCAGGGCCAGCATCTTCAG |
| **sgCCL17_2** | GACGTGGTACCAGACATCTG |
| **sgCCL17_3** | TCTCTGCAGCACATCCACGC |
| **sgCCL17_4** | GCAGCTCGAGGGACCAATGT |
| **sgCCL17_5** | GCAGCACTCCCGGCCCACAT |
| **sgCCL18_1** | TGGTGCAGACGAGGACAAGG |
| **sgCCL18_2** | GACAGACTCACCTTGTGCAC |
| **sgCCL18_3** | TCTGCCAGGAGGTATAGACG |
| **sgCCL18_4** | ACACCTGGCTTGGGGCACTG |
| **sgCCL18_5** | CTTTTGTGGAATCTGCCAGG |

| | |
|---|---|
| **sgCCL22_1** | AGAAGTGTTTCACCACGCGC |
| **sgCCL22_2** | GGACCTCAGACTCCTGCCCG |
| **sgCCL22_3** | GCGCCACAGCAAGGAGGACG |
| **sgCCL22_4** | TGTGGCGCTTCAAGCAACTG |
| **sgCCL22_5** | GACGGTAACGGACGTAATCA |
| **sgCD40LG_1** | CATGCTGATGGGCAGTCCAG |
| **sgCD40LG_2** | ACGATACAGAGATGCAACAC |
| **sgCD40LG_3** | CCAGTTTGAAGGCTTTGTGA |
| **sgCD40LG_4** | AAAGTGCTGACCCAATCATC |
| **sgCD40LG_5** | GCAGTCCAGTGGCCGCAGAT |
| **sgCD80_1** | AAGCCATGGGCCACACACGG |
| **sgCD80_2** | GAAGTGGCAACGCTGTCCTG |
| **sgCD80_3** | GAGCCAGGATCACAATGGAG |
| **sgCD80_4** | GCTCTGCGCCCATCTGACGA |
| **sgCD80_5** | AGGCTCTGGAAAACCTCCAG |
| **sgCD82_1** | GACTTCGCAGGAACAGGGGT |
| **sgCD82_2** | CCAGAGTGGCAACCACCCTG |
| **sgCD82_3** | ACACAGGCCAGTCCTCAGGG |
| **sgCD82_4** | GGATCCACACCCCGAAGCCC |
| **sgCD82_5** | CCTGCAGCTGAAGCAGGAGA |
| **sgCD86_1** | GAATGAAACAGACAAGCTGA |
| **sgCD86_2** | GGAAATGGAAGAAGAAGAAG |
| **sgCD86_3** | AGAGTGAACAGACCAAGAAA |
| **sgCD86_4** | CAGGAACCAACACAATGGAG |
| **sgCD86_5** | GGAAGTACAGCTGTAATCCA |
| **sgCDC25A_1** | GTGCTGGAGCTACACAGGGA |
| **sgCDC25A_2** | AAAGAGATAGCAGTGAACCA |
| **sgCDC25A_3** | CCAGAGGCTTGCCATGCACG |
| **sgCDC25A_4** | AGAGTCAACTAATCCAGAGA |
| **sgCDC25A_5** | GAACTTGCACATGGAAGAAG |
| **sgCDH2_1** | GCAGTGTACAGAATCAGTGG |
| **sgCDH2_2** | GAGATAGTACCATTGAGAAG |
| **sgCDH2_3** | GACTGGTGACATCATCACAG |
| **sgCDH2_4** | TGGCCTTTCAAACACAGCCA |
| **sgCDH2_5** | GCAGAGCAGGATGGCAATGA |
| **sgCDK12_1** | GAAGCAATTGTGGGAAGAGG |
| **sgCDK12_2** | GGCAAGAAGGACGGGAGTGG |
| **sgCDK12_3** | GGAGACTGATGACTATGGGA |
| **sgCDK12_4** | GTCCTGGAAGTACCTCGAGA |

| | |
|---|---|
| **sgCDK12_5** | GCTGGCTGAGAAGGAGGCAG |
| **sgCDK2AP1_1** | GCCATGGAGAGGCTGAAGCG |
| **sgCDK2AP1_2** | TGGGTGTAGCCTAGGGACGG |
| **sgCDK2AP1_3** | GCTGGCCATCATTGAAGAGC |
| **sgCDK2AP1_4** | GGAGCAAGAGTGCCATGGAG |
| **sgCDK2AP1_5** | GCAGCTCCGCGTATTTGCTT |
| **sgCDK6_1** | GCCCGCGACTTGAAGAACGG |
| **sgCDK6_2** | ATGCGTGGCGGAGATCGGGG |
| **sgCDK6_3** | CGGGGAGGGCGCCTATGGGA |
| **sgCDK6_4** | GCCGCTCTCCACCATCCGCG |
| **sgCDK6_5** | AACATTCTGGTGACCAGCAG |
| **sgCDYL2_1** | AATGGGAGTATCTTATCCGA |
| **sgCDYL2_2** | GAAGCGAATTAACCCTCCCC |
| **sgCDYL2_3** | GCCCAGGAGGCCTGCAGCAG |
| **sgCDYL2_4** | GACCTCCTGGCTGAACGTGG |
| **sgCDYL2_5** | GAAAGGCTACGGGAGCACCG |
| **sgCEACAM1_1** | GAAATGCAGAAAACATGCCA |
| **sgCEACAM1_2** | CAACAACTCCAACCCTGTGG |
| **sgCEACAM1_3** | GGTACACGCACTCTGTGAAG |
| **sgCEACAM1_4** | GCATGGATTCAGTAGTGAGC |
| **sgCEACAM1_5** | GAGAGGCTGAGGTTTGCCCC |
| **sgCENPV_1** | CGGGAGCAAGAGCCAGGCGG |
| **sgCENPV_2** | AACGCCAGCTCAGGCGGCGG |
| **sgCENPV_3** | GGCCTGGTGAAGCACACAGG |
| **sgCENPV_4** | ACTCCACGATCAAACCCCGG |
| **sgCENPV_5** | GGCGCTGGGTGCCAAGGCAG |
| **sgCEP170B_1** | TGTGCAGCACCGAGTCCCGG |
| **sgCEP170B_2** | CAGGAGGGCTGCCCATACAG |
| **sgCEP170B_3** | GGTAGGGCTCTCCCTGGCGC |
| **sgCEP170B_4** | GCAGCCCTGAGGCTCAGCAG |
| **sgCEP170B_5** | GGCACACGAGATGCCCACGA |
| **sgCHST15_1** | GAAGGGAACGAAAACTGGGG |
| **sgCHST15_2** | AAGTCCTGCGTCAGAAACGG |
| **sgCHST15_3** | GCAGCAGGTCAACTGCCAAG |
| **sgCHST15_4** | TCATCACCTTTCCATTACGG |
| **sgCHST15_5** | AGAACTCCTCGTACCAACAG |
| **sgCITED2_1** | GGGCCGGGGACTGTGAACGG |
| **sgCITED2_2** | GGAGTTGTTAAACCTGGCCG |
| **sgCITED2_3** | GCGGGCAGCAGCAACAGCGG |

| | |
|---|---|
| **sgCITED2_4** | ACTGCCCCATGCCCATGCGG |
| **sgCITED2_5** | CGAGCACATACACTACGGCG |
| **sgCLEC4F_1** | GAACCAATGCTGAGATCCAG |
| **sgCLEC4F_2** | GAGCTCCACGTGCTAAGCAG |
| **sgCLEC4F_3** | GAACACCAAAGCTCTAACCA |
| **sgCLEC4F_4** | GTGAAGTAATGACCACATGG |
| **sgCLEC4F_5** | GGGGTAGCCTGAACGAGCCT |
| **sgCSF1_1** | GAGACAGACCAACAACAGCA |
| **sgCSF1_2** | GGCGAGCAGGAGTATCACCG |
| **sgCSF1_3** | GACACCTGCAGTCTCTGCAG |
| **sgCSF1_4** | GTAGAACAAGAGGCCTCCGA |
| **sgCSF1_5** | GAGTTCCTGCAGCTGCACAA |
| **sgCSF2_1** | CCTGCTTGTACAGCTCCAGG |
| **sgCSF2_2** | GCATGTGAATGCCATCCAGG |
| **sgCSF2_3** | GAGACGCCGGGCCTCCTGGA |
| **sgCSF2_4** | GCTTGGTGAGGCTGCCCCGC |
| **sgCSF2_5** | GACCTGCCTACAGACCCGCC |
| **sgCSF2RB_1** | GAGGGCCAGGAGCACAGCGA |
| **sgCSF2RB_2** | CGTAGATGCCACAGAAGCGG |
| **sgCSF2RB_3** | GTGACGACAAAACTCTGGCA |
| **sgCSF2RB_4** | GAGCCACTGGTTGTCCCCAG |
| **sgCSF2RB_5** | AGGGGATCTGGAGTTTGAGG |
| **sgCSTB_1** | GATGTGCTGGGTCTCGGCGG |
| **sgCSTB_2** | CGTGTCATTCAAGAGCCAGG |
| **sgCSTB_3** | GGGGACAAACTACTTCATCA |
| **sgCSTB_4** | GTCGCCGCCAAGATGATGTG |
| **sgCSTB_5** | AATGACACGGCCTTAAACAC |
| **sgCTPS1_1** | GCTGATGAAATGGAAAGAGA |
| **sgCTPS1_2** | GACATGCCAGAACACAACCC |
| **sgCTPS1_3** | GAATCCTCACCTCAAATCGG |
| **sgCTPS1_4** | GAGGAGGCCAAAGTATGGTG |
| **sgCTPS1_5** | AAGTACATTCTGGTTACTGG |
| **sgCTU2_1** | GCAGAGCACCAAAGCACCGA |
| **sgCTU2_2** | GGTCTGAGCCCCAAAAGCCG |
| **sgCTU2_3** | GAATGGGCTTCACTTCGGCC |
| **sgCTU2_4** | TCTAAGGCCACCACATGCCA |
| **sgCTU2_5** | TCCTGGCCCGACTCAAGGGG |
| **sgCX3CR1_1** | AAAGCCGATGAAGAAGAAGG |
| **sgCX3CR1_2** | GCACGGGCCAGATTTCCTGG |

| | |
|---|---|
| **sgCX3CR1_3** | CAACCGTCTCAGTCACACTG |
| **sgCX3CR1_4** | CCAGGCATTTCCCATACAGG |
| **sgCX3CR1_5** | CAGATGAGGAGAAATCAACG |
| **sgCXCL10_1** | ATGTCTGAATCCAGAATCGA |
| **sgCXCL10_2** | GATAAGGCAGCAAATCAGAA |
| **sgCXCL10_3** | ACTCACATGATCTCAACACG |
| **sgCXCL10_4** | TGAAAGCAGTTAGCAAGGAA |
| **sgCXCL10_5** | CAGAATGGCAGTTTGATTCA |
| **sgCXCL11_1** | AAACATGAGTGTGAAGGGCA |
| **sgCXCL11_2** | GAAGGGCATGGCTATAGCCT |
| **sgCXCL11_3** | AGCACACAATATCACAGCCA |
| **sgCXCL11_4** | GTTACTTGGGTACATTATGG |
| **sgCXCL11_5** | GGCTTCCCCATGTTCAAAAG |
| **sgCXCL9_1** | GGAACCCCAGTAGTGAGAAA |
| **sgCXCL9_2** | CCTGCATCAGCACCAACCAA |
| **sgCXCL9_3** | GGCAGTGCTACACTGAAGAA |
| **sgCXCL9_4** | ATTTTCTCGCAGGAAGGGCT |
| **sgCXCL9_5** | CCAGCAAGATGATGCCCAAG |
| **sgCYFIP1_1** | GTACGACGAAATTGAGGCCG |
| **sgCYFIP1_2** | ACGAGACGCTGCTGAAGCAG |
| **sgCYFIP1_3** | GCTGAGCCGGTACCTGACGC |
| **sgCYFIP1_4** | TGAGCTCCCAGAAGACGTGC |
| **sgCYFIP1_5** | GCAGTACGTGAAGACGCTGA |
| **sgCYP1B1_1** | GCCACTGATCGGAAACGCGG |
| **sgCYP1B1_2** | GCCGCCGGACACCACACGGA |
| **sgCYP1B1_3** | ACGGCGCCTTCCTCGACCCG |
| **sgCYP1B1_4** | GTAGCGGCAGCCGAAACACA |
| **sgCYP1B1_5** | TGAGCAGCTCACGGAACTCG |
| **sgDFNA5_1** | GCAAACTCTCGAAAGACCAG |
| **sgDFNA5_2** | GGAGGATGGGAATGTCACCA |
| **sgDFNA5_3** | GTATAACTCAATGACACCGT |
| **sgDFNA5_4** | GTGGGGTGCAGCTTACAGGG |
| **sgDFNA5_5** | TGGGATCTCCAGCACCACGT |
| **sgDMD_1** | GAAGGAGGCCCAGCCTGCGC |
| **sgDMD_2** | CAGGACGGGCAGCCACACCA |
| **sgDMD_3** | AGATGTTCGAGACTTTGCCA |
| **sgDMD_4** | GGTGGGCATCATTTCAGGAG |
| **sgDMD_5** | GCAATGAGCTCAGCATCCCG |
| **sgDNAJC6_1** | GACATGGTAAAAGGAGGTGC |

| | |
|---|---|
| **sgDNAJC6_2** | AAAGGTTACTCACCAAGCAG |
| **sgDNAJC6_3** | GTAGGTCACACATATAGCCC |
| **sgDNAJC6_4** | AGTGTCACCTGAAATAGCTG |
| **sgDNAJC6_5** | GAATTGGTGGGAGGAGCCGC |
| **sgDOCK7_1** | GAAGGTGATGGAGATGCAGG |
| **sgDOCK7_2** | GAAATGGTACGGCGAAGCCG |
| **sgDOCK7_3** | GAATCTGATGAAGCTCCAGA |
| **sgDOCK7_4** | GAAGTAGAAATCAGTACTGG |
| **sgDOCK7_5** | GCAGAAGAAACAGAAGGTGA |
| **sgDPP3_1** | GATGTACTGGGTGTCCGCCA |
| **sgDPP3_2** | GCAGCACCCAGAAGAAGTCA |
| **sgDPP3_3** | AGAGAACATAAGCTCCCCGC |
| **sgDPP3_4** | GGAAGATGCCAAATTGGCCC |
| **sgDPP3_5** | AGGGATGCCGGAGCCAGCGA |
| **sgDYRK3_1** | TGCAGAGTACATTTGAACAG |
| **sgDYRK3_2** | TTTAGTCTTCTGGGTGGAGG |
| **sgDYRK3_3** | GGAGGGTATGATGATGCAGA |
| **sgDYRK3_4** | GGTGCTGAAAATTATTGGCA |
| **sgDYRK3_5** | GGCATCCAAAGATTGCAAGA |
| **sgECE1_1** | GGAAACCCGAAAATCAGCCA |
| **sgECE1_2** | GAACTGGGTGAAGAAGAACG |
| **sgECE1_3** | CATGGAGCTCAAGATGGAGC |
| **sgECE1_4** | GAGCTCCATGGACCCCACAG |
| **sgECE1_5** | CTGGGGAAGCTGCTGGGCGG |
| **sgEGLN3_1** | GGACAACTTCCTGGGCGAGG |
| **sgEGLN3_2** | CAAGCAGCTGCACTGCACCG |
| **sgEGLN3_3** | GAGGGACAGGAGGAAGCTGA |
| **sgEGLN3_4** | GCAAATACTACGTCAAGGAG |
| **sgEGLN3_5** | GATGCAGCGACCATCACCGT |
| **sgEMP1_1** | GCTGGAACACGAAGACCAGG |
| **sgEMP1_2** | GCATAATAACAGTAGCGATG |
| **sgEMP1_3** | GCTGTCACTGCAGCTAATGT |
| **sgEMP1_4** | AGATGCCCTCAAGACAGTGC |
| **sgEMP1_5** | CCAGCTCTTCACCATGGAGA |
| **sgENO2_1** | GAGGCCCTGGAGCTGAGGGA |
| **sgENO2_2** | CGGGAGATCCTGGACTCCCG |
| **sgENO2_3** | GATGAGGGCTGGCGCGATGG |
| **sgENO2_4** | GAAACTGGACAACCTGATGC |
| **sgENO2_5** | ACACGGCCAGAGACACACCC |

| | |
|---|---|
| **sgENPP6_1** | GAAACACCAGCAGCTTCCGG |
| **sgENPP6_2** | GAGATGCAGGCCCGTAGTGG |
| **sgENPP6_3** | GAATGGATCAGAACCTCTGT |
| **sgENPP6_4** | GAATGACCTGCAGCAAGTGA |
| **sgENPP6_5** | GCAGCATGGCAGTGAAGCTT |
| **sgETF1_1** | TGGCTCCCAGGACTGAAAGG |
| **sgETF1_2** | TATACAACCAGACCATTTGG |
| **sgETF1_3** | GAATGAATTTCACGTTGGAG |
| **sgETF1_4** | AGAGCTTGGAGGCGGCCCGC |
| **sgETF1_5** | CAGTGCTGCCGACAGGAACG |
| **sgETS2_1** | CCTGTGGCTAACAGTTACAG |
| **sgETS2_2** | GTTAATCCAATGAGGAACGG |
| **sgETS2_3** | GAAGTGGACCTATTCAGCTG |
| **sgETS2_4** | GCTCTGGCCATACCTCATCG |
| **sgETS2_5** | GTTGCTCACTCCACAGCCAG |
| **sgEXT2_1** | CCAAAACTGAAACATGCCAG |
| **sgEXT2_2** | CAAGGAGACAGCACAAGCGA |
| **sgEXT2_3** | GACAGGAGGAAGTATTGCCG |
| **sgEXT2_4** | GACCCACCTGAGATGACAGG |
| **sgEXT2_5** | ACTGTGGTGGGATCAGCGGG |
| **sgFAM129B_1** | GAACCTCTTCCAGCACCAGG |
| **sgFAM129B_2** | GGTGGAAAACACGTACGAGG |
| **sgFAM129B_3** | CGTGCCCCACAACTACGGGC |
| **sgFAM129B_4** | GGCACTGCCCGACTTTGCCG |
| **sgFAM129B_5** | GACGCGCATAAGGATGCCAG |
| **sgFAM13A_1** | GTATCTCCAGGAAGAAGAGG |
| **sgFAM13A_2** | GAGCAGGATGAAGTTCGACA |
| **sgFAM13A_3** | TGACAAAGCAGCCAATCCGG |
| **sgFAM13A_4** | GAAGCTCCAGGAGAAGCGAG |
| **sgFAM13A_5** | CCTGCTCCTTGGACTGGCGG |
| **sgFAM162A_1** | GGAGCTCCGGGACTTTCCTG |
| **sgFAM162A_2** | GTGGCCATGGGGAGCCTCAG |
| **sgFAM162A_3** | AGTGCCTTTACACAAACCTA |
| **sgFAM162A_4** | TCTAATGATTGCCCTGACGG |
| **sgFAM162A_5** | GATGCATCCTACCACCGTCA |
| **sgFAM219A_1** | GTAGGGGCTCACCTCAGCAG |
| **sgFAM219A_2** | GTCTCCGTCAGAGATGGAGG |
| **sgFAM219A_3** | GGAGACTGTGACGCCCGGGA |
| **sgFAM219A_4** | GTTCTCCTGCAGAGAAGCAG |

| | |
|---|---|
| **sgFAM219A_5** | CTGCATCTCCGAGTGCGCGG |
| **sgFBXO22_1** | GCGGCTCCTCCGTAGACCCG |
| **sgFBXO22_2** | GAGTAACCTGGCGGAGGTGG |
| **sgFBXO22_3** | GATCTCCGCAGGCCTGGCGG |
| **sgFBXO22_4** | TCAGCGTGTGCCGCTTATGG |
| **sgFBXO22_5** | GAGTGTCGTGGCCATAAGAG |
| **sgFMOD_1** | CCAGCAAGATCAGTGACCGG |
| **sgFMOD_2** | GCAGTCGCGGGGATCTGGAG |
| **sgFMOD_3** | GAAGTACACATACTTCATGC |
| **sgFMOD_4** | GGAGGGGCTGGAGAACCTCA |
| **sgFMOD_5** | GTACAGCTGCTCAAGAGCTG |
| **sgFOSL1_1** | CGGCCAAGTGCAGGAACCGG |
| **sgFOSL1_2** | GGCCTTCGACGTACCCCTGG |
| **sgFOSL1_3** | CTGCTGGGCTGCCTGCGCTG |
| **sgFOSL1_4** | GCCCGGATGACTCCTGGCCG |
| **sgFOSL1_5** | GCCGCCGAGTAAGGCGCGAG |
| **sgFOXO1_1** | GGAGAGCGAGGACTTCCCGC |
| **sgFOXO1_2** | GGACTGGCTAAACTCCGGCC |
| **sgFOXO1_3** | GAGCAACCTGAGCTTGCTGG |
| **sgFOXO1_4** | GGTTGCCCCACGCGTTGCGG |
| **sgFOXO1_5** | GGGTCGATCTCCACCACCTG |
| **sgFOXO3_1** | GTGGGTACGCACCTTCCAGC |
| **sgFOXO3_2** | GCTCCTTGAGGACTCGGCCC |
| **sgFOXO3_3** | CACTTCGAGCGGAGAGAGCG |
| **sgFOXO3_4** | CAGAGTGAGCCGTTTGTCCG |
| **sgFOXO3_5** | CCTGCCATATCAGTCAGCCG |
| **sgGART_1** | CGAGTACTTATAATTGGCAG |
| **sgGART_2** | AGTGTGGTCACTGATTGAGA |
| **sgGART_3** | GAGATGGTGGCCCTAACACA |
| **sgGART_4** | TTATCCTGGAGACTACACCA |
| **sgGART_5** | GCTTTGGTTGTGAAGGCCAG |
| **sgGAS7_1** | CAGTAGTTCAAACCCAGCCA |
| **sgGAS7_2** | GAAGGGCAAACAAATGCAGA |
| **sgGAS7_3** | GTGTTGCAGCTCCTTGGGAG |
| **sgGAS7_4** | GGAGATGAAGACCCAGCAGC |
| **sgGAS7_5** | CAAGCTGAGCAACAAGACAG |
| **sgGBE1_1** | GAGTTGGAACAGCATTGCCA |
| **sgGBE1_2** | GTACTTGAAGCCCTACGCCG |
| **sgGBE1_3** | TGTGCTAGATTCAGATGCAG |

| | |
|---|---|
| **sgGBE1_4** | GTAGTCCTCGGGCCGAGCCG |
| **sgGBE1_5** | GGCGTCCACAGATGTGCTGA |
| **sgGJB2_1** | GAGTGAATTTAAGGACATCG |
| **sgGJB2_2** | ATAGAAGACGTACATGAAGG |
| **sgGJB2_3** | GGAGGTGTGGGGAGATGAGC |
| **sgGJB2_4** | TGCATGGCCACTAGGAGCGC |
| **sgGJB2_5** | GAAGAAGAGGAAGTTCATCA |
| **sgGJB6_1** | GCAGGTGGTACCCATTGTAA |
| **sgGJB6_2** | ACACTCCACCAGCATCGGGA |
| **sgGJB6_3** | GGTGGCTGCCCAGGAAGTGT |
| **sgGJB6_4** | TGCATGGCCACCAGCAGCGC |
| **sgGJB6_5** | GGATAGAGGGGTCGCTGTGG |
| **sgGNE_1** | GATGTCAGTGCCAAAAGATG |
| **sgGNE_2** | GGGAGAAGATGAGGCAGCCA |
| **sgGNE_3** | GGAACACCTGTGATCAACCT |
| **sgGNE_4** | GGTACTGTTTACCAAACTGA |
| **sgGNE_5** | GATTGTTCCTGTGGAAGCCA |
| **sgGNG8_1** | CAACATGGCCAAGATTGCCG |
| **sgGNG8_2** | GGTGAACATCGACCGCATGA |
| **sgGNG8_3** | TGCCGAGGCCCGCAAGACGG |
| **sgGNG8_4** | GACGGTGGAACAGCTGAAGC |
| **sgGNG8_5** | CCCGACCCAGCACCTTCATG |
| **sgGOLIM4_1** | GACGCAGCTGCGGAAAGCCG |
| **sgGOLIM4_2** | GGAGCACAGAAAGGCCCTGG |
| **sgGOLIM4_3** | GCAAGAGAATACAGAAGTGG |
| **sgGOLIM4_4** | GTAGAGAAAGCCGAACACGA |
| **sgGOLIM4_5** | AGAGCCAATCCAACAAGAAG |
| **sgGORASP2_1** | ACACCAAGTAACCTGTGGGG |
| **sgGORASP2_2** | GGTTGAGGTTGAGGTTGGGG |
| **sgGORASP2_3** | AAAGCGTCGAGATCCCGGGC |
| **sgGORASP2_4** | ACACCAAATTCTGCATGGGG |
| **sgGORASP2_5** | CATTGGTGGCACAGAAGTGA |
| **sgGOT1_1** | GGAACCTCGGCAAAGACTGA |
| **sgGOT1_2** | GCTCACTGCCGACTTCAGGG |
| **sgGOT1_3** | CTACTGGCAAAACCCAGGGA |
| **sgGOT1_4** | GATGTGGTTCCAGGTCCCAG |
| **sgGOT1_5** | CCGCAAGGTCAACCTGGGAG |
| **sgGPM6B_1** | GAGGATGGTGGCCACCAGGG |
| **sgGPM6B_2** | AGGCCACCCCGGAGAAGCAG |

| | |
|---|---|
| **sgGPM6B_3** | ACAGATCTGCTCCACACCCG |
| **sgGPM6B_4** | GCCGCAGAATAAGGCCACCC |
| **sgGPM6B_5** | TGTGGCTCTCGCAGGCACCG |
| **sgGRINA_1** | CCAGCATGGAAACTACCAGG |
| **sgGRINA_2** | AAAAGTGAACACAGACACCG |
| **sgGRINA_3** | GGAAGTCCCCACAACAGCTG |
| **sgGRINA_4** | TAAGGGGCCCCAGGGTAGGG |
| **sgGRINA_5** | GGTACCCTGGCTGACCGTAG |
| **sgGTF2E2_1** | CAGAATTTCAGAAACTGTGG |
| **sgGTF2E2_2** | TGATAGACACGGCATCAGCG |
| **sgGTF2E2_3** | GAAACAATGGCTAATGACTG |
| **sgGTF2E2_4** | GGAATCTGGACCAAAGAAAG |
| **sgGTF2E2_5** | AAAGCTTTGTCAGGAAGCTC |
| **sgGZMB_1** | AAGTCTCTGAAGAGGTGCGG |
| **sgGZMB_2** | TAAGATAAGCCATGTAGGGG |
| **sgGZMB_3** | GCCACACTGCATGTCTGCCC |
| **sgGZMB_4** | ACACTCACACACACTACAAG |
| **sgGZMB_5** | AGAGGTGAAGATGACAGTGC |
| **sgHBEGF_1** | AAGAGCTTCAGCACCACCGA |
| **sgHBEGF_2** | GGTCCGTGGATACAGTGGGA |
| **sgHBEGF_3** | AAGGGGCACTTACATGCAGG |
| **sgHBEGF_4** | TAGGGGTAGCAGCTGGTCCG |
| **sgHBEGF_5** | GAAAGTCCGTGACTTGCAAG |
| **sgHEG1_1** | GATGGGAACAGAGAGGGCGA |
| **sgHEG1_2** | GAGCTGGCAGACCTCAGCAG |
| **sgHEG1_3** | GCAGATGCTAGAGGAAGGAG |
| **sgHEG1_4** | GCCGCTTTTCAAACAAAGAG |
| **sgHEG1_5** | CGGGCTCCAAGCCGAAGTGG |
| **sgHILPDA_1** | GTCCAGGAGGTCCCAGGCGA |
| **sgHILPDA_2** | GAACCTCTACCTGTTAGGTG |
| **sgHILPDA_3** | AGCCAACACAGAGCCCACCA |
| **sgHILPDA_4** | AGAGTGATGGAGTCCCTAGA |
| **sgHILPDA_5** | GTTGGCTAGTTGGCTTCTGG |
| **sgHMOX1_1** | AGGGCCTCTGACAAATCCTG |
| **sgHMOX1_2** | TGTGAAGCGGCTCCACGAGG |
| **sgHMOX1_3** | AAAGATTGCCCAGAAAGCCC |
| **sgHMOX1_4** | GGACCTGCCCAGCTCTGGCG |
| **sgHMOX1_5** | GGCACTGGCAATGTTGGGGA |
| **sgHOXB9_1** | GGGGACAATAAAATTTGCGA |

| | |
|---|---|
| **sgHOXB9_2** | GTACAGTTTGGAAACTTCGG |
| **sgHOXB9_3** | GGAACTCCAGGTGCTCCGCG |
| **sgHOXB9_4** | TTGGGCTGGAAGCTGCACGA |
| **sgHOXB9_5** | CCTGGGGCTGGATGTAAGGG |
| **sgHSF5_1** | GCAGCTGTGGCTGGTCGCGG |
| **sgHSF5_2** | ACAACTTCCCCGCCAAGCTG |
| **sgHSF5_3** | GATGAAGCTGGTGAAGCTGG |
| **sgHSF5_4** | CTGGTGCTGCAGTGGCGCGG |
| **sgHSF5_5** | CCAGGCGCCACAGCTTGGCG |
| **sgHSPG2_1** | GGAGCGAGTGAAATTCACCA |
| **sgHSPG2_2** | GAAGCCTGGATGGACTCCCG |
| **sgHSPG2_3** | TGATCTGGTCCCGGAAGCGG |
| **sgHSPG2_4** | GAATTGCCAGCACAACACGG |
| **sgHSPG2_5** | GGGCTGCCAGCATCACACGG |
| **sgIFNG_1** | AAAGAGTGTGGAGACCATCA |
| **sgIFNG_2** | CCAGAGCATCCAAAAGAGTG |
| **sgIFNG_3** | TGCAGGTCATTCAGATGTAG |
| **sgIFNG_4** | GACATTCATGTCTTCCTTGA |
| **sgIFNG_5** | ACATGAACTCATCCAAGTGA |
| **sgIGLL1_1** | GGTCCAGCCTGAGGAGCCGG |
| **sgIGLL1_2** | GTTGGGGCCTGGCTCACCAG |
| **sgIGLL1_3** | GCCCCAACCTCAGGCAGCGC |
| **sgIGLL1_4** | GCTGCTGCTGGGTCTGGCCG |
| **sgIGLL1_5** | GGTCTGGCCGTGGTAACCCA |
| **sgIL17F_1** | TACGTGACATGGAAACGCGC |
| **sgIL17F_2** | GCTTGCCTTTCTGAGTGAGG |
| **sgIL17F_3** | GCACTTACGTGTAATTCCAG |
| **sgIL17F_4** | GACCCTGCATGGCCCAGCCA |
| **sgIL17F_5** | AGTTGCCCGCCTGTGCCAGG |
| **sgIL27RA_1** | TGGAAAAGCACCCACAACAG |
| **sgIL27RA_2** | GCAGCGGCCACAGCCAGAAA |
| **sgIL27RA_3** | GCTCTGGAGGTGTAACTCGG |
| **sgIL27RA_4** | GACATGGTGAGCTGTTCCCG |
| **sgIL27RA_5** | CAAATCTTGGATCTCAACAG |
| **sgIL4I1_1** | GAAGCTGCGCAACTATGTGG |
| **sgIL4I1_2** | GGAGGCAGATAACAGGATCG |
| **sgIL4I1_3** | AGGGTGATTGTGGTTGGCGC |
| **sgIL4I1_4** | ACCGGGACCAGAACACGGGC |
| **sgIL4I1_5** | ATCCTCCACAAGCTCTGCCA |

| | |
|---|---|
| **sgILDR2_1** | GTCGCGGAAGAACTTCGCCA |
| **sgILDR2_2** | GAAGTGGCAGCGAAGCACAG |
| **sgILDR2_3** | GGACACCGGAGACAGAGGGA |
| **sgILDR2_4** | GAGTCACTTGGTGCCAAGGG |
| **sgILDR2_5** | GGTGTCATGGGAGGCAGCAG |
| **sgIPO11_1** | GCAGCAGTGGAAGAAACAGG |
| **sgIPO11_2** | GTGGTGATTCCACAGAGAGC |
| **sgIPO11_3** | CAAGGGTATTATAGAAGGGG |
| **sgIPO11_4** | CTGAGGAGCAGTTGAAGCAG |
| **sgIPO11_5** | GGAACCTCATAAGAATATGG |
| **sgIRF4_1** | CACGCGGGGCATGAACCTGG |
| **sgIRF4_2** | GGCATGAGCGCGGTGAGCTG |
| **sgIRF4_3** | GCAACGGGAAGCTCCGCCAG |
| **sgIRF4_4** | GCAGGACTACAACCGCGAGG |
| **sgIRF4_5** | GCGTTGTCATGGTGTAGGGG |
| **sgIRGQ_1** | GCAGCTCTGCTGAACAGCGC |
| **sgIRGQ_2** | GCGGTCACGTCACCCTGCGG |
| **sgIRGQ_3** | GCGGACTTCCCCAAGCCCGG |
| **sgIRGQ_4** | TCTGTGAATCCCCAGGACGG |
| **sgIRGQ_5** | GCGAACTGCGGCAGCAGCGA |
| **sgJUNB_1** | GGCGAGCTTGAGAGACGCGC |
| **sgJUNB_2** | GGGGTAAAAGTACTGTCCCG |
| **sgJUNB_3** | GGAGGAACCGCAGACCGTGC |
| **sgJUNB_4** | CAGGCGTTCCAGCTCCGAAG |
| **sgJUNB_5** | GGTAGCGCCCAGGGACACGT |
| **sgKDM3A_1** | CATACTCCAAACCCACACCG |
| **sgKDM3A_2** | GGAGTGCTGGAAACAAGGGC |
| **sgKDM3A_3** | TCTTCAAGTCAACTGTGAGG |
| **sgKDM3A_4** | GGTGCATTTGAAACATCCGA |
| **sgKDM3A_5** | TGGCTGGCCGACCTAACCAG |
| **sgKIAA1211_1** | GTTGGGAGGACGCGGAGCGG |
| **sgKIAA1211_2** | GGAGGAAGGAAGATGCGCGG |
| **sgKIAA1211_3** | GGAGGAGCTCAAAAGGCAGG |
| **sgKIAA1211_4** | GGAAGAGCTGGAACAGCAGG |
| **sgKIAA1211_5** | GGAGCTGCGGTGGCAGGAGG |
| **sgKIAA1217_1** | GCTGTGGAAGGGCTGGCCGG |
| **sgKIAA1217_2** | GGAGGGCTGGATGACCACAG |
| **sgKIAA1217_3** | GAGGAGGTGGTGACAGGCGG |
| **sgKIAA1217_4** | TGGTTGACCTCACCACAGGA |

| | |
|---|---|
| **sgKIAA1217_5** | GAATTTGGAATCGCATGGGG |
| **sgKIT_1** | GAAAGAAGACAACGACACGC |
| **sgKIT_2** | CCAAACCTGAACACCAGCAG |
| **sgKIT_3** | GGGTGCTGGAGCTTTCGGGA |
| **sgKIT_4** | ATATTCTGTAATGACCAGGG |
| **sgKIT_5** | GCTGAGCTTTTCTTACCAGG |
| **sgKLC2_1** | GAACTGCTGAAAGATGGCAG |
| **sgKLC2_2** | AGGGACTGGAGACTCTGCGT |
| **sgKLC2_3** | GCAGCCCCTAGCCCAGGAGG |
| **sgKLC2_4** | GGATCAGAACAAGTACAAGG |
| **sgKLC2_5** | GAACCAGGGCAAAGCTGAGG |
| **sgKLF2_1** | GTAATAGAACGCAGGCGGCG |
| **sgKLF2_2** | TCAGACACCAGGCCACCCGC |
| **sgKLF2_3** | CAGGCGGCGTGAGGAGACCG |
| **sgKLF2_4** | CTGGAGGCCAAGCCAAAGCG |
| **sgKLF2_5** | GTTTGCGGGGCCAAGAGCGG |
| **sgLARP4_1** | GATGCCACCCCAGTAACTCA |
| **sgLARP4_2** | CGACATGTTGCTTTTCGTGG |
| **sgLARP4_3** | ACATCAGGTGCTCATCCTGA |
| **sgLARP4_4** | AATTTGGACAGTTGCCAACA |
| **sgLARP4_5** | GTCAGACACAGATGCACAAC |
| **sgLEF1_1** | GCAGTCAAATAAAGTGCCCG |
| **sgLEF1_2** | GTAGGGATATCAGGAGCTGG |
| **sgLEF1_3** | TGGGTGTCTTACCGTCATCG |
| **sgLEF1_4** | GTACCCGGAATAACTCGAGT |
| **sgLEF1_5** | AGAGGATGGACCGCATGGGA |
| **sgLRP5_1** | GTACGGCTGGTGGACGCCGG |
| **sgLRP5_2** | GAAGAGGCGGACCCTCCTGG |
| **sgLRP5_3** | GGAGGGACTTGGACAACCCG |
| **sgLRP5_4** | GGAAAACTGGAAGTCCACTG |
| **sgLRP5_5** | GTACTGGACAGACGTGAGCG |
| **sgLRP8_1** | GACGGGGAGAAGGACTGCGA |
| **sgLRP8_2** | GCAGCGCCAGAAGCCGGAGA |
| **sgLRP8_3** | GCAGCTCCAGCATCTTGCGG |
| **sgLRP8_4** | GCGGCTGATCCGCTGCTCGG |
| **sgLRP8_5** | GCACCTTGGCCGCCGAGCAG |
| **sgLRRC59_1** | GGATCTCCTCAACAACAAGC |
| **sgLRRC59_2** | AGGAGCGGGAGAGGCAGCGG |
| **sgLRRC59_3** | AGGCGGAAGAGAAGGAGCGC |

| | |
|---|---|
| **sgLRRC59_4** | ACCAAGGCCGGTAGCAAGGG |
| **sgLRRC59_5** | GGACCTCATTCAGGTCGCTG |
| **sgMAK16_1** | GGAGAGACTGAAACAAGATA |
| **sgMAK16_2** | CAAAGCCCTGGAACAACAGG |
| **sgMAK16_3** | GCATACTGACTATTTGCCAG |
| **sgMAK16_4** | ACCCGTTCCCAGAGACGCCG |
| **sgMAK16_5** | TGAGTAAGAAGGTGGAGCGT |
| **sgMAN2A2_1** | GGAGCTGACAGCCAACGCAG |
| **sgMAN2A2_2** | GCTACAGGCAGAGCGAGCGC |
| **sgMAN2A2_3** | GATATGACAAGCCCCAGGAG |
| **sgMAN2A2_4** | GAGTCCTGGAAGCCCACCTG |
| **sgMAN2A2_5** | GCTGTAGCTGCAGCACGCCC |
| **sgMAP1B_1** | GAAGGAAGGCAAGGCCGCAG |
| **sgMAP1B_2** | CGGCAGCATCGCCAACCCGG |
| **sgMAP1B_3** | CAGCAAGTTCTACTTGCTGG |
| **sgMAP1B_4** | GACTGGTACCAACAAAGACA |
| **sgMAP1B_5** | GTACAACATCCTGGAAGGGT |
| **sgMAPK11_1** | GAAGCGCATCATGGAAGTGG |
| **sgMAPK11_2** | GACATGTCGGGCCCTCGCGC |
| **sgMAPK11_3** | GGAGCTGAACAAGACCGTGT |
| **sgMAPK11_4** | GCAGAACGTACCGGGAGCTG |
| **sgMAPK11_5** | CAGGTCGGCGCCCATCAGGG |
| **sgMED27_1** | ATAGTGGAGCAGGGCAGTGG |
| **sgMED27_2** | GCCCCAGATGCCGGATGTCG |
| **sgMED27_3** | GCTGCGCTCCAGCGTGAGCA |
| **sgMED27_4** | AGCCACTTACCATGAAGGAT |
| **sgMED27_5** | GTCCTGGAAGTGCGCAATAA |
| **sgMETTL2A_1** | GCTGCTCCACAGGATGCAGA |
| **sgMETTL2A_2** | TAAAACACAGACACTTCCTG |
| **sgMETTL2A_3** | GCAGGCTTCTGAAACCTGGC |
| **sgMETTL2A_4** | CCTGCAGTCGGCGATCCACC |
| **sgMETTL2A_5** | AGAGTGATTACCAGGCATTG |
| **sgMICAL2_1** | GAAGCAGCTGGATCAGGAGA |
| **sgMICAL2_2** | GAAGTCGTGCACGAACACCA |
| **sgMICAL2_3** | GGGAGCCAAAGTGGTCGTGG |
| **sgMICAL2_4** | TGAGCTCAGGCCGGAAGCGG |
| **sgMICAL2_5** | AGAGCTCGTAGAACTTGGAG |
| **sgMIEF1_1** | AAGAAGGATGACAATGGCAT |
| **sgMIEF1_2** | GCTGAACAGGGACATGAAGA |

| | |
|---|---|
| **sgMIEF1_3** | GCAGGCGCTGGTGAGCGCAA |
| **sgMIEF1_4** | AGGTGGAGGAGTCTGTGGGA |
| **sgMIEF1_5** | GCCAGTGGCCAGGAAGGGCC |
| **sgMLEC_1** | CGAAGTGCCCATCAAAGAGG |
| **sgMLEC_2** | GCTGGTCTTGAAATTTGCAG |
| **sgMLEC_3** | AGGGGAAGCTGAGTGTCCAG |
| **sgMLEC_4** | GCACGGGATCCACTTCCGCA |
| **sgMLEC_5** | GGACCCTTTGGAAGGCCGGG |
| **sgMOB3A_1** | GACTTGCTTCAGGAAGGGGT |
| **sgMOB3A_2** | GCTCAAACTTGCGCTTGGGG |
| **sgMOB3A_3** | CGAGCTGCACAAGAAGGCGC |
| **sgMOB3A_4** | TGAACGCCGGGCTGGACCTG |
| **sgMOB3A_5** | GCAACTGCACGGCCAGCCGC |
| **sgMRC2_1** | GGACGATCGGAGCTGCACGG |
| **sgMRC2_2** | GAGGCTGGTATTGCAAGCCG |
| **sgMRC2_3** | GGGAGACCCACTTCCAGCGC |
| **sgMRC2_4** | GCCGCGAGGATGGTCACCTG |
| **sgMRC2_5** | ACTACGGCAAAGACGAGCGC |
| **sgMT1E_1** | ACCAGTGGCGCAAGAGCAGT |
| **sgMT1E_2** | CTGCTGTTCCTGCTGCCCCG |
| **sgMT1E_3** | CTGCGTCTGCAAAGGGGCAT |
| **sgMT1E_4** | GCCCAGGGCTGCGTCTGCAA |
| **sgMT1E_5** | GACGCAGCCCTGGGCACACT |
| **sgMT1H_1** | GCGTTCCTTACCAGCCTCGC |
| **sgMT1H_2** | GGACCCCAACTGCTCCTGCG |
| **sgMT1H_3** | CAGCCTCGCAGGAGCAGTTG |
| **sgMT1H_4** | GCACACTTGGCACAGCCCAG |
| **sgMT1H_5** | GCACTTGCAGGAGCCGGCGC |
| **sgMT2A_1** | CCACTTACTTTTCTTGCAGG |
| **sgMT2A_2** | CAGCCTCTTACCGGCGGCGC |
| **sgMT2A_3** | GCATTTGCAGGAGCCGGCGC |
| **sgMT2A_4** | GCAGGAGCAGTTGGGATCCA |
| **sgMT2A_5** | ACCGGCGGCGCAGGAGCAGT |
| **sgMXI1_1** | CGGGCACAGAAACACAGCAG |
| **sgMXI1_2** | GTGGCGACTGGAACAGCTGC |
| **sgMXI1_3** | GCTGCAGGGTCCTCAGGAGA |
| **sgMXI1_4** | TGCCCGGCTCAACCTCCGTG |
| **sgMXI1_5** | GGAGATGGAACGAATACGAA |
| **sgMYB_1** | AGAAATACGGTCCGAAACGT |

| | |
|---|---|
| **sgMYB_2** | GAAAACAATGTAGGGAGAGG |
| **sgMYB_3** | GACGCATTGTAGAATTCCAG |
| **sgMYB_4** | GGAATAGTCGTTGTTAACAG |
| **sgMYB_5** | GAACTTACTAGCATCTTCAG |
| **sgNCEH1_1** | GCTGTGTACAGCAATGGCTG |
| **sgNCEH1_2** | ATTTCTGGTGACAGTGCTGG |
| **sgNCEH1_3** | GGAAATCTGGCTGCTGCCCT |
| **sgNCEH1_4** | CAGATACAGGCTAGTTCCAA |
| **sgNCEH1_5** | CAGAAGTATATGGTTGATCC |
| **sgNCOA2_1** | GCACTTCAGGCCCTCAGCGA |
| **sgNCOA2_2** | GGTCTGGCGAACCTCCGAGG |
| **sgNCOA2_3** | AGAACCACCAAACCTGCCCA |
| **sgNCOA2_4** | GCAGGGGAAAACTGACTGGG |
| **sgNCOA2_5** | GGAATGCAAGCTTCCTGCAG |
| **sgNID1_1** | TACAAGGCTCTGAGAAGGGG |
| **sgNID1_2** | AGAGTGCAGGGACTACGCCA |
| **sgNID1_3** | GCAATCTGGTCATTAAGCAG |
| **sgNID1_4** | GCGGACTTGGACACGACCGA |
| **sgNID1_5** | GCAGCAGAGTGTGTCCACAG |
| **sgNID2_1** | GGGAACAGAAGAGTGGGCAG |
| **sgNID2_2** | CAGAAGCCCAGCTCCACCAG |
| **sgNID2_3** | GCAGTGGCAGCAGAAGCCAG |
| **sgNID2_4** | AACCTGCTGGAGCACTACGG |
| **sgNID2_5** | GCGGCCCGCAACATTAGCAA |
| **sgNOL6_1** | GGCCTGGTGGAAGTCAGCCA |
| **sgNOL6_2** | GCAGCCAGTGAAGCTCAGCA |
| **sgNOL6_3** | AAGGCTGTCAGAGAAGAAGA |
| **sgNOL6_4** | AGGGCACCCTCACAACCCGC |
| **sgNOL6_5** | GCAACAGTGAGGGTTTCAGG |
| **sgNOTCH2_1** | GGGGACTGGGGCAGAAGCAG |
| **sgNOTCH2_2** | GTGCCGTGTAGAGTTCACAG |
| **sgNOTCH2_3** | GGAGGGGTGAGAGCCAGCGA |
| **sgNOTCH2_4** | GCCTGGAGTACAGGAGGCGA |
| **sgNOTCH2_5** | CACAGGGCACAAGCAAGAGA |
| **sgNR4A1_1** | GAACAGACAGCCTGAAGGGG |
| **sgNR4A1_2** | GAAGGTGCTGAAGCTGGGCA |
| **sgNR4A1_3** | GGATGTGGAGGAGGCCGAGG |
| **sgNR4A1_4** | GAAGTCCTCGAACTTGAAGG |
| **sgNR4A1_5** | ATAGTAGTCAGAGCCACTGG |

| | |
|---|---|
| **sgNR4A2_1** | GAAGCTGGGGAGAGAAGTGG |
| **sgNR4A2_2** | TAAAGGAGAAGAGGGAGAGG |
| **sgNR4A2_3** | AATGGGGTTGGGCACAGCGA |
| **sgNR4A2_4** | GAAGTGCCTGGCTGTTGGGA |
| **sgNR4A2_5** | GGCACTGATCAGACTCACCG |
| **sgNR4A3_1** | GGACTGCTTGAAGTACATGG |
| **sgNR4A3_2** | GAAGAGCGGGAAGCGCGCGC |
| **sgNR4A3_3** | ATACAGCTCGGAATACACCA |
| **sgNR4A3_4** | GATGCTGGGCAGGGACGTGG |
| **sgNR4A3_5** | CCTGCGTGTACCAAATGCAG |
| **sgNRG4_1** | CAAGTCGTTTTGCCTGAATG |
| **sgNRG4_2** | CACTCACCTACAAAATGGGC |
| **sgNRG4_3** | CTAGATCACGAAGAGCCCTG |
| **sgNRG4_4** | GACTTGTGACTGGGACCACA |
| **sgNRG4_5** | AGGTGCGTTGAAAACTATAC |
| **sgNRIP3_1** | GCAGCGCCGGATGAAGCAGG |
| **sgNRIP3_2** | GGAGACCGACATGCGGGAGG |
| **sgNRIP3_3** | GGGGCTAAAGAAGTCTGAGG |
| **sgNRIP3_4** | CTGCTGGGCAGTCCAGGCGG |
| **sgNRIP3_5** | GTCCAGGGGCAGCAGGTCGG |
| **sgNUDT15_1** | GGCAGTTTCCAACTCCCTGG |
| **sgNUDT15_2** | GCAACGCGGATGCTTGCAGC |
| **sgNUDT15_3** | GTTGCGTCCTCCTGGGGAAG |
| **sgNUDT15_4** | GGGAAGAGGAAAGGCTCGGT |
| **sgNUDT15_5** | CGAACTCCAGATGACCTCCA |
| **sgNUP93_1** | ACAAAACTGAGGATTACCTG |
| **sgNUP93_2** | GAGCTGAGCACTCTGTCCAG |
| **sgNUP93_3** | TGAGGAAATAGAAGTACTGG |
| **sgNUP93_4** | AGGGATGAGAAAGATAGTCA |
| **sgNUP93_5** | GGAAAGTGTGGAAGAGAGAG |
| **sgOGFOD1_1** | GATGAATGGGAAGCGGCCAG |
| **sgOGFOD1_2** | CATGTGGAATGGAGCAGCCG |
| **sgOGFOD1_3** | CAGGTTTTATGAGAAAGCTG |
| **sgOGFOD1_4** | TGAGCCAGAGAATAATCAGA |
| **sgOGFOD1_5** | GGGATCACACAGTGAAGAAA |
| **sgOLIG3_1** | GAGAGAGCGGACAGGTGCGG |
| **sgOLIG3_2** | ATATGATGCAGAAGATGCCC |
| **sgOLIG3_3** | GCACGACCTGAACCTAGCCA |
| **sgOLIG3_4** | GATGGCGGGAAGTGAGGCGG |

| | |
|---|---|
| **sgOLIG3_5** | GCCTTGAGTAGCGAGTGGGG |
| **sgOPA1_1** | CAAGTGGAATGACTTTGCGG |
| **sgOPA1_2** | ATACGCAAGATCATCTGCCA |
| **sgOPA1_3** | AGGAACTTTTAACACCACAG |
| **sgOPA1_4** | GAAGAATCGGACCCAAGAAC |
| **sgOPA1_5** | TAAAAGGCCCTGGACTACAG |
| **sgOSBPL6_1** | AGAGTATTCACATACTGGAG |
| **sgOSBPL6_2** | CTGAATGGGGAGCTTACAGG |
| **sgOSBPL6_3** | GGGGTGGTGATAGATCAGGA |
| **sgOSBPL6_4** | CCAGAAACCAGACAAACATG |
| **sgOSBPL6_5** | GGTCCATGGGAGCATAGATG |
| **sgOSM_1** | GGAGCTGGCCAAGGAGCACG |
| **sgOSM_2** | GAAGCAGACAGATCTCATGC |
| **sgOSM_3** | GCAGGACACCAGCAGACTCC |
| **sgOSM_4** | ATAGGGGTCCAGGAGTCTGC |
| **sgOSM_5** | TCCAAGCATGGCGAGCATGG |
| **sgP4HA2_1** | GACCCAGGCACAATTTCCAG |
| **sgP4HA2_2** | TACCAAGGGAGAGCAGGCGG |
| **sgP4HA2_3** | GCTGAGCTAGCAACCCCAGA |
| **sgP4HA2_4** | GCAGCCCACAAGCACAGGGC |
| **sgP4HA2_5** | GAAAGAGTACATCCTTGTGG |
| **sgPDGFA_1** | CCAGCCTCTCGATCACCTCG |
| **sgPDGFA_2** | GCTGGAGGTCCCGGATGCTG |
| **sgPDGFA_3** | GGGCGAGGTATCCGCAGCCG |
| **sgPDGFA_4** | AGCGTTTCACCTCCACGCAC |
| **sgPDGFA_5** | GGATTATCGGGAAGAGGACA |
| **sgPDIA5_1** | GAGCGAGGAGACCTTTGCAG |
| **sgPDIA5_2** | GTACTCAAAATAGCAGATGG |
| **sgPDIA5_3** | GACAAGCGTGTTGCACCTGG |
| **sgPDIA5_4** | GAAGAAGAAGAAACACACCT |
| **sgPDIA5_5** | AGGGGCGTAGAACATGACCA |
| **sgPFKFB4_1** | GGAACCGCCGGACGTCACGG |
| **sgPFKFB4_2** | GTACCGGTACCCTAAAGGGG |
| **sgPFKFB4_3** | GCACTGACCTCCCTGCCCCG |
| **sgPFKFB4_4** | GAGGACAATCCAGACGGCTG |
| **sgPFKFB4_5** | TGGGTGTGCCCTATGAACAG |
| **sgPGAP1_1** | GAACTTCAATGAAGAACTGG |
| **sgPGAP1_2** | TAACTTACCACACAATGGAG |
| **sgPGAP1_3** | GAAGACATCCCACAGCCCCA |

| | |
|---|---|
| **sgPGAP1_4** | GGCATCACAGGAGCAACATG |
| **sgPGAP1_5** | CATTGCACTTAGAAAAGCAG |
| **sgPGK1_1** | TAAGGTGCTCAACAACATGG |
| **sgPGK1_2** | GCATGGGCACACCATCAGGC |
| **sgPGK1_3** | GGAGCTGAACTACTTTGCAA |
| **sgPGK1_4** | AGAGTTCTTCATACCCGCCC |
| **sgPGK1_5** | GCTTTCTAACAAGCTGACGC |
| **sgPIP5K1A_1** | AAGGCTCAACCTACAAACGG |
| **sgPIP5K1A_2** | AGAACCTCAACCAGAACCCT |
| **sgPIP5K1A_3** | ATCTGGAATCAAGAGACCCA |
| **sgPIP5K1A_4** | GAAACTCACCTCAGATGCCA |
| **sgPIP5K1A_5** | GTAAAACATCAGGACGACCA |
| **sgPIP5K1B_1** | GCTGGGAATAGGATACACAG |
| **sgPIP5K1B_2** | GAAGCAGCTTCAAACAGTGA |
| **sgPIP5K1B_3** | GGAAATGAATTCCCAACAGA |
| **sgPIP5K1B_4** | GGTTCTCTACTCAACAGCCA |
| **sgPIP5K1B_5** | GGTCCAGGGAAATCTGGAGA |
| **sgPLEKHG1_1** | GCAGAGGGACAACCCCGCCA |
| **sgPLEKHG1_2** | GCTGTCGCGGGAAGAGGCCG |
| **sgPLEKHG1_3** | GTGGAGAGTGGACTCAAACG |
| **sgPLEKHG1_4** | GGACCAAACAAAACTTCCCC |
| **sgPLEKHG1_5** | GAAGCGGAAACACGAGCACG |
| **sgPLK3_1** | GGATCTGGTCAATAGAGGGG |
| **sgPLK3_2** | GCGCGGAGACAGGAAACCGG |
| **sgPLK3_3** | GGTCCGCGCAAGGCACTCGG |
| **sgPLK3_4** | GGTCAGCCGTCTCAAAGGGA |
| **sgPLK3_5** | CCTGAAGACAGCTCACCCCG |
| **sgPLOD2_1** | GACAGGGATTCCAGGGGTGG |
| **sgPLOD2_2** | CAAGGAGAAGAATGGAGAGG |
| **sgPLOD2_3** | AACAATATGTCCAGAAACCG |
| **sgPLOD2_4** | CCAAAAGGCAAACCACAAAG |
| **sgPLOD2_5** | AAACGCTATCTGAATTCAGG |
| **sgPMAIP1_1** | GTACCTGCTGGAGCCCGCGC |
| **sgPMAIP1_2** | ACGCTCAACCGAGCCCCGCG |
| **sgPMAIP1_3** | GGCGGAGATGCCTGGGAAGA |
| **sgPMAIP1_4** | CGGCACCGGCGGAGATGCCT |
| **sgPMAIP1_5** | TCAGATTCAGAAGTTTCTGC |
| **sgPNO1_1** | GGGCCGCATGGACACAGAGG |
| **sgPNO1_2** | AGCCCGTCCCCACAGAGGGG |

| | |
|---|---|
| **sgPNO1_3** | GAGGTCATCCAACCTGATGA |
| **sgPNO1_4** | AATGGAAACGCAGAGCGCCA |
| **sgPNO1_5** | AGAGTGGGAAAGAAGAAACA |
| **sgPNPO_1** | CATGCGCAAGAGTTACCGCG |
| **sgPNPO_2** | GTAGAAGACAAGGGAAGCAA |
| **sgPNPO_3** | GAAGAAACTGCCTGAGGAGG |
| **sgPNPO_4** | GTGTGGTCGCAGTGCTGCCA |
| **sgPNPO_5** | GGTAGCCAGACACATGGCAT |
| **sgPOLR3K_1** | CGGGAACGGGCTGATCGTGG |
| **sgPOLR3K_2** | ACGTGTTGCAGGAGAAGCGG |
| **sgPOLR3K_3** | GTACCCAAAACTGAAAGAAG |
| **sgPOLR3K_4** | GCCGGGGCAGAACAGCAGCA |
| **sgPOLR3K_5** | GGGTGATGTTGTGCACGTAG |
| **sgPOMGNT1_1** | GCAGAGGCAGTGGTCCCCGG |
| **sgPOMGNT1_2** | TCTACGTCCAGGACCCGCCG |
| **sgPOMGNT1_3** | CAGCAAAGTATATGTGGCAG |
| **sgPOMGNT1_4** | GGGGTATGAACACACGGCTG |
| **sgPOMGNT1_5** | CGGATGCCTGAACAACGCCG |
| **sgPPRC1_1** | GGTAGCCCAAGACTGAGGGG |
| **sgPPRC1_2** | GAAGCCTGAAGGAGTTACGG |
| **sgPPRC1_3** | GGAGAAGTGACAGGAGGCGG |
| **sgPPRC1_4** | GACTTCCCCAACCACTGCCG |
| **sgPPRC1_5** | CCAGGTGCTGCTGCATGAGG |
| **sgPRDX1_1** | CCACAGAGCGGCCAACAGGG |
| **sgPRDX1_2** | ACCATCTGGCATAACAGCTG |
| **sgPRDX1_3** | GGGGTCTTAAAGGCTGATGA |
| **sgPRDX1_4** | TCAAAGTCTCATCCACAGAG |
| **sgPRDX1_5** | TTCACTGACAAACATGGGGA |
| **sgPRG2_1** | GGAGACACCAGAGCAGGAGA |
| **sgPRG2_2** | GGGCCACGCAGTGACCACCG |
| **sgPRG2_3** | AGAGGACACAGTAAAAGTGG |
| **sgPRG2_4** | GACCCCTTGCAGGGAGCTGG |
| **sgPRG2_5** | GAAGATGCCTCCAAGAAAGA |
| **sgPROSER1_1** | GGAGTGGAGGAACCATGTGG |
| **sgPROSER1_2** | GAAAAGTGAGGGAACAGCAG |
| **sgPROSER1_3** | GAAACCCAGGGATACCCGGG |
| **sgPROSER1_4** | GAATGTACTAACGAAGGCAA |
| **sgPROSER1_5** | GGAGTATATGGAGGTGGAGG |
| **sgPSMB2_1** | TCAGGAAGGCACCATAGCCG |

| | |
|---|---|
| **sgPSMB2_2** | AGGCGACAAGAACATAGTCG |
| **sgPSMB2_3** | CATCATAGCCAGCCAGGAGG |
| **sgPSMB2_4** | GAGGATACTGAGAGTCAGGA |
| **sgPSMB2_5** | CCTTAGGAAATGTCTGGAGG |
| **sgPSMD3_1** | GGATGTGGAGATGAAAGAGG |
| **sgPSMD3_2** | GGCCAGCATCAACCACGAGA |
| **sgPSMD3_3** | AAGGCGAAACCGCCGCCCGG |
| **sgPSMD3_4** | AAAGAGGAGGCAGCGACGGG |
| **sgPSMD3_5** | GGCAGACGGCAAGACGGCGG |
| **sgPSME3_1** | GAAAAGCAACCAGCAGCTGG |
| **sgPSME3_2** | AGCCCACTTATAAGAAGCGA |
| **sgPSME3_3** | GCTCTGATATGATGAGCCGA |
| **sgPSME3_4** | ATTGGGCATCACAAACACCT |
| **sgPSME3_5** | CAATAATGTCCACCAGCTGC |
| **sgQTRTD1_1** | AAGATACTTCTCCATCGGAG |
| **sgQTRTD1_2** | GCGGCTGCAGGAAGAGTCAG |
| **sgQTRTD1_3** | AGTCACTGCAGAGCTGCCGG |
| **sgQTRTD1_4** | CGAGTGTATTGAAAGAGGAG |
| **sgQTRTD1_5** | GTGTCTCTTCAGGATTCGGC |
| **sgRAB19_1** | GTGAGGGATGGACTCGAACG |
| **sgRAB19_2** | GCACTTCTCCAGCTCAGCCA |
| **sgRAB19_3** | ACAGCAGAACACGATTGGAG |
| **sgRAB19_4** | GGTGAGGTCATAGGCGATGA |
| **sgRAB19_5** | GTCAATATCAAGGGAACGCA |
| **sgRAB3GAP1_1** | CGAATAGCAATACTAACTGG |
| **sgRAB3GAP1_2** | ACATGGCCTCATCTGACCGA |
| **sgRAB3GAP1_3** | GCATTTGAGGAAGAAGGCAA |
| **sgRAB3GAP1_4** | CAAGAGAGTGGCAAGAAAGG |
| **sgRAB3GAP1_5** | TGAGAACTTGAAGTCAGCAA |
| **sgRAB3GAP2_1** | GAAGTAGCAGAACTGGACAA |
| **sgRAB3GAP2_2** | AGGAAAGAGGAAGTCCCGGG |
| **sgRAB3GAP2_3** | GGTCCAGTCAGGACGCCCAG |
| **sgRAB3GAP2_4** | TGACTGGACCTGCATTGTGG |
| **sgRAB3GAP2_5** | GCATGGAGAAAGCTCCACTG |
| **sgRALB_1** | AAGGTGATCATGGTTGGCAG |
| **sgRALB_2** | ATAGATATTCTGGACACCGC |
| **sgRALB_3** | AGTCTGACCTAGAGGAGCGG |
| **sgRALB_4** | GCGGAGGCAGGTGCCTGTGG |
| **sgRALB_5** | GCTTCAGTTCATGTATGACG |

| | |
|---|---|
| **sgRARG_1** | CCAGATCCAGCTGCACGCGG |
| **sgRARG_2** | CCAGCTGGGCAAGTATACCA |
| **sgRARG_3** | GCAGCTCCTGCCCCTGGAGA |
| **sgRARG_4** | GCTACAGAAGTGCTTCGAAG |
| **sgRARG_5** | CAAGAAGAAGAAAGAGGTGA |
| **sgRDH10_1** | GGGCGAACTCCAGCGCGAAG |
| **sgRDH10_2** | GTCACGATGTACATGAGGCG |
| **sgRDH10_3** | CGTCGCGATGAACATCGTGG |
| **sgRDH10_4** | GCCCAAGGAGAAGAGCGTGG |
| **sgRDH10_5** | CATCTACCGCGACCTGGAGG |
| **sgRNASEH1_1** | GCACTGGAGAGGCTTACCCA |
| **sgRNASEH1_2** | TCAACATGGACAAGAATCGG |
| **sgRNASEH1_3** | TCAGGGCATGAAAATCAACA |
| **sgRNASEH1_4** | AAAGCACATGAAGCCGAGCG |
| **sgRNASEH1_5** | TCACCAGGTCAGAAAGACCC |
| **sgRRP12_1** | GAGGTGATTCGCTCCCAGGG |
| **sgRRP12_2** | GGAGACGCCCATGGAAGAAG |
| **sgRRP12_3** | GCCACTCAGGAAGGTACCCG |
| **sgRRP12_4** | GAAGCGCTGTACTTTGCTGA |
| **sgRRP12_5** | GTCTTGCTTCCGCAGAAGGG |
| **sgS1PR4_1** | GAGCAGGCTGAAGGTGGAGG |
| **sgS1PR4_2** | GGGAGGCCATGAACGCCACG |
| **sgS1PR4_3** | CGAGTCCTGCCAACAGCTGG |
| **sgS1PR4_4** | GGAGAACTTGCTGGTGCTGG |
| **sgS1PR4_5** | GGGGCAGAAGGCTGGAGCAG |
| **sgSAMD4A_1** | CGCGTGTGAACTGCCCGGGG |
| **sgSAMD4A_2** | GCAATCCAACTCCCTCCCGA |
| **sgSAMD4A_3** | ACGGCTGGCTCTGTGGGCGG |
| **sgSAMD4A_4** | CCTGTGGTTTGCCAACCCCG |
| **sgSAMD4A_5** | GAGGGGATCTGGTACTGGCG |
| **sgSEMA7A_1** | GACCTTGTCAATGTCACCGA |
| **sgSEMA7A_2** | GCACATGAAAGGTCTCCCCG |
| **sgSEMA7A_3** | CAGGGGCACTATCCACAAGG |
| **sgSEMA7A_4** | TCTGGATGGCAGCCGCGCGG |
| **sgSEMA7A_5** | ATGGTGGCTTTGATGAACTG |
| **sgSETBP1_1** | CAAGGAGGCAAACTTCACAG |
| **sgSETBP1_2** | AGAGCGCATGGAGCCAGAGG |
| **sgSETBP1_3** | GGGCGAGTCAGACTTCCTGC |
| **sgSETBP1_4** | ACCTGGGAAGGGGATCCCGG |

| | |
|---|---|
| **sgSETBP1_5** | GAGGAGGATGAACTAGGCTC |
| **sgSETD7_1** | CGAGATGGTGGAGGAGGCGG |
| **sgSETD7_2** | GGATAGCGACGACGAGATGG |
| **sgSETD7_3** | GACGATGACGGATTACCGCA |
| **sgSETD7_4** | TTCACGGAGAAAAGAACGGA |
| **sgSETD7_5** | GATGATGCCTTGCAGGGCCA |
| **sgSH2B3_1** | TGAGTTGCACGCCGTAGCGG |
| **sgSH2B3_2** | GGAGAGCCGGACATCACAGG |
| **sgSH2B3_3** | GAACTCGCTCCAGCCCCGCG |
| **sgSH2B3_4** | GGCGGCGGAAGATGTGGCGG |
| **sgSH2B3_5** | GAAGAGCTCCAGCACGCGGT |
| **sgSHC4_1** | GTGCTGAGTGGACTACAGGG |
| **sgSHC4_2** | GATCAAGGTGCACAGTGGGG |
| **sgSHC4_3** | GCATATTGATAGCCATGCCG |
| **sgSHC4_4** | TCTGAAACACCACCTACTGG |
| **sgSHC4_5** | GCGGATCAAAGTTCAAGCCA |
| **sgSKAP1_1** | GATCTCCTCAGGGAGGGCGG |
| **sgSKAP1_2** | GATCGGAGTGGCAGAAGCGA |
| **sgSKAP1_3** | GGGGATAAAAGAGCCTACAG |
| **sgSKAP1_4** | AGTGGCACTCGACGAAAAGG |
| **sgSKAP1_5** | GAACCTCAGCGCTGTTGCAA |
| **sgSLC16A9_1** | GGACAATCAGAAGATTGGAG |
| **sgSLC16A9_2** | AGACCTGTCACAATCTTCAG |
| **sgSLC16A9_3** | CAGTATTTTGACGATCGCCG |
| **sgSLC16A9_4** | GCTGGACTTGGAAATAGCCT |
| **sgSLC16A9_5** | GAGACACAGAGACTGCAGAC |
| **sgSLC25A15_1** | GAAGCCACGGAAGCCCACCT |
| **sgSLC25A15_2** | GGCAGAGGCGAAGGAACCGG |
| **sgSLC25A15_3** | GCAGAGCACCAGTGCAGCAA |
| **sgSLC25A15_4** | ACGTTCCCTGACCTGTACCG |
| **sgSLC25A15_5** | GCTTCTGCCAGCAGGTGGTG |
| **sgSLC25A33_1** | GGTTTCCCTTACAGGTGTGG |
| **sgSLC25A33_2** | GAACTCACCCGCCGGCGAAG |
| **sgSLC25A33_3** | CAGGTCGATCTTGGAGAAAG |
| **sgSLC25A33_4** | ACCGTGTGGATAAGCAATGC |
| **sgSLC25A33_5** | ACAGTCCTCTATAAAAGGCA |
| **sgSLC26A4_1** | CAGGTCATGGCAGCGCCAGG |
| **sgSLC26A4_2** | GTTGTGAAGCCACCAACCAA |
| **sgSLC26A4_3** | AGCATAAGCCACCACAGCGA |

| | |
|---|---|
| **sgSLC26A4_4** | CGTGCGGGAAAGAGCAGTGG |
| **sgSLC26A4_5** | GCCGTCCAGGAGAGCACTGG |
| **sgSLC35F2_1** | GATGGAGGCAGACTCGCCAG |
| **sgSLC35F2_2** | GTACAGGCAAAACATACACA |
| **sgSLC35F2_3** | CAAGGACAACATCTGACCCA |
| **sgSLC35F2_4** | AGTGATCCACTTCATCGCCG |
| **sgSLC35F2_5** | ACCCAACAGACAGACAGCCA |
| **sgSLC37A3_1** | GGGATCGGTTGAATTTGCGA |
| **sgSLC37A3_2** | AATGTTGCCCACCGAAGCAC |
| **sgSLC37A3_3** | GGAACACTACAACATGATGA |
| **sgSLC37A3_4** | GAGTCGCTTTCTCTGCACTG |
| **sgSLC37A3_5** | AAATGGTATCCAGTGTGCCG |
| **sgSLC39A14_1** | GGAGAATGAGCAGACGGAGG |
| **sgSLC39A14_2** | GGCCCACTCCCACATCCAGG |
| **sgSLC39A14_3** | TGGATGATCACTCTGAGCGA |
| **sgSLC39A14_4** | CATGTGGGAACTCCTCACAG |
| **sgSLC39A14_5** | TATTAGATCCTGCAGGAAGG |
| **sgSLC7A5_1** | GATACTCACAGATGCCCTGG |
| **sgSLC7A5_2** | GGGCTTCGTCCAGATCGGGA |
| **sgSLC7A5_3** | GGACGACAGCATCTGCTCGG |
| **sgSLC7A5_4** | GTAGCGGCTCACCACGGCCA |
| **sgSLC7A5_5** | GCTGTGGGTGGATCATGGAG |
| **sgSMOX_1** | CGTGTGGTGGTGATCGGCGC |
| **sgSMOX_2** | GAGGCTTCCAGCCACATCGG |
| **sgSMOX_3** | GGTGGCTCCCAGCTCAAAGG |
| **sgSMOX_4** | GCGGATGACCCTCTCAGTCG |
| **sgSMOX_5** | ACCGATGGGGAACGCAGCGT |
| **sgSRM_1** | GGACCACGGACTCCACGGAG |
| **sgSRM_2** | GGTGCTTCACCACCTCCCGC |
| **sgSRM_3** | GGTCCAGTGTGAGATCGACG |
| **sgSRM_4** | GCAGCCGCTGACACAGCAGC |
| **sgSRM_5** | GGCGGTACCTGCGGAAGACG |
| **sgSRPK1_1** | CCTGCAGATCCGGAAGGCGG |
| **sgSRPK1_2** | GGCCCGAAAGAAAAGGACCA |
| **sgSRPK1_3** | TGAATGAGCAGTACATTCGG |
| **sgSRPK1_4** | GAAGCGAATGCAGGAAATTG |
| **sgSRPK1_5** | TGAGGAAATGGAGAAAGAGT |
| **sgSRXN1_1** | GCTGCGTGCAGGAGGAACGC |
| **sgSRXN1_2** | GGGTCCAACACGGACGGCAG |

| | |
|---|---|
| **sgSRXN1_3** | CGAGGCTCTGCACCTTGGCG |
| **sgSRXN1_4** | GGAGGAACGCTGGGCAGGGC |
| **sgSRXN1_5** | GCAGGGCGGCAGCATCCACT |
| **sgSTIP1_1** | AGGGCGACTACAATAAGTGC |
| **sgSTIP1_2** | GATCTGCTGCACCTCAGGGT |
| **sgSTIP1_3** | GGCAACAAGGCCCTGAGCGT |
| **sgSTIP1_4** | GAAAGGCACTTACTGAAGGT |
| **sgSTIP1_5** | TACACGGAAAGCCGCTGCGC |
| **sgTAF13_1** | AGATGAGGAAGAAGACCCCA |
| **sgTAF13_2** | GAAATTGGAGGAGGTGCAGA |
| **sgTAF13_3** | GAATCCTTATACTGAGTCAG |
| **sgTAF13_4** | CGATGTATGATGTATGGCTT |
| **sgTAF13_5** | GGACCCAAGGAAGTTTGCCA |
| **sgTBC1D24_1** | GGAGTGGCTGAGTGGGGCGG |
| **sgTBC1D24_2** | GAATGTCCCGGATCAGGCGC |
| **sgTBC1D24_3** | GGTGAACAAGTACTGCCAGG |
| **sgTBC1D24_4** | GGAGATCCAGCTCCTGCAGA |
| **sgTBC1D24_5** | GCAGGGAGGAGAACAGCAGA |
| **sgTCF7_1** | GATGTCCGCAGGTGCAGGGG |
| **sgTCF7_2** | GAGAGGTCAGGGGTCTGCAG |
| **sgTCF7_3** | GTAGAAGCCAGAGAGGTCAG |
| **sgTCF7_4** | GAGCTGCCCCATGCTGCCTG |
| **sgTCF7_5** | GCAGCTCCTGCTTCCCTGAG |
| **sgTMEM33_1** | TGAGTGAATACAACAGGTAG |
| **sgTMEM33_2** | GATGACCAATAAACTGGACA |
| **sgTMEM33_3** | GTTCGGGGTCGTATCTGCCA |
| **sgTMEM33_4** | GCTGTCCTCTAACAAAGCCT |
| **sgTMEM33_5** | GCTCTGCTTAACTGGAAGTG |
| **sgTNF_1** | GGCGCTCCCCAAGAAGACAG |
| **sgTNF_2** | GGACGTGGAGCTGGCCGAGG |
| **sgTNF_3** | GCTGAGGAACAAGCACCGCC |
| **sgTNF_4** | TCAGGAAGGAGAAGAGGCTG |
| **sgTNF_5** | CAGCAGGCAGAAGAGCGTGG |
| **sgTNFRSF10B_1** | ACAACGGGGACAGAACGCCC |
| **sgTNFRSF10B_2** | CCAGAGCTCACAACGACCTG |
| **sgTNFRSF10B_3** | GCAGATGCTCTGACTCCCCG |
| **sgTNFRSF10B_4** | CCTACCGCCATGGAACAACG |
| **sgTNFRSF10B_5** | AAGGCACGGCCCAGGACCCA |
| **sgTNFRSF4_1** | GCAGAGACGAGGATGTGCGT |

| | |
|---|---|
| **sgTNFRSF4_2** | GGTAGGTGTCCCCGACACAG |
| **sgTNFRSF4_3** | GACGGGGCTCCACTGTGTCG |
| **sgTNFRSF4_4** | GCCGGGCTTCTACAACGACG |
| **sgTNFRSF4_5** | ACTGAGGTTACACCACGTGC |
| **sgTNFRSF9_1** | GGAGCGTGAGGAAGAACAGC |
| **sgTNFRSF9_2** | GTTGAGGACCAGCAACAGAG |
| **sgTNFRSF9_3** | AGGTGTTTTCAGGACCAGGA |
| **sgTNFRSF9_4** | TGTGCTTGTGAATGGGACGA |
| **sgTNFRSF9_5** | GGAGAGAGGTCGGCTGGAGA |
| **sgTOX2_1** | GCAGGAGGGATGGCTCCGGG |
| **sgTOX2_2** | GTAGGCAGGGAGCAGACCGT |
| **sgTOX2_3** | GCAGCTGGCCCGACAGCAGG |
| **sgTOX2_4** | GGAGATGGTCCACTCGGAAG |
| **sgTOX2_5** | GATGAGCTGGGACTGGCTGA |
| **sgTOX3_1** | GCAGGGTCTGACTCTGGAGG |
| **sgTOX3_2** | GATGGCGTGCTTCATAGCAG |
| **sgTOX3_3** | GATGGACCGCATGATCAGGG |
| **sgTOX3_4** | ATTGATGGAGCTGGAAGGGG |
| **sgTOX3_5** | AGAATACCTGAAGGCCCTGG |
| **sgTP53I3_1** | GCAACGCTGAAATTCACCAA |
| **sgTP53I3_2** | GGAAAACCTCTACGTGAAGG |
| **sgTP53I3_3** | GAGGTGGCCAAGCCGAGCCC |
| **sgTP53I3_4** | GAAGATCGGGGACACAGCCA |
| **sgTP53I3_5** | GCTCCTCATGCCTATCCCAG |
| **sgTRAF4_1** | GGCAGGTATCGCAGAAACGG |
| **sgTRAF4_2** | AGTGTGCAGGTAGATCACGG |
| **sgTRAF4_3** | GAATGGGCAGAGCACCAGGG |
| **sgTRAF4_4** | GCAGAAACGGTGGCCGCAGG |
| **sgTRAF4_5** | GCTTCCTCTGGACTATGCCA |
| **sgTRAM2_1** | CAAGGGTGAGGATGAAGAGG |
| **sgTRAM2_2** | GTAGAAGCACCAAATCACCG |
| **sgTRAM2_3** | CATGCGGGTAGTCTTCCCAG |
| **sgTRAM2_4** | GCAGCAAGATCAGGCCCAGG |
| **sgTRAM2_5** | GCTTCATCCACTCCCAGCTG |
| **sgTRIP10_1** | GGACTGCCGGGAGGCAGAGA |
| **sgTRIP10_2** | GAAGGTGGCTGCAAATGCTG |
| **sgTRIP10_3** | GTCGCAGCCGCACCAAGCGC |
| **sgTRIP10_4** | CATCACAGGACACCAAGGAG |
| **sgTRIP10_5** | GTTTGATGAGGATTTCGAGG |

| | |
|---|---|
| **sgTSPAN15_1** | GTCACTCACCCAGAACACGG |
| **sgTSPAN15_2** | GGAGCTCATTGGTGGCGTGG |
| **sgTSPAN15_3** | GGACAACTACACCATCATGG |
| **sgTSPAN15_4** | TAAGTGAAAACTTGAGCCAG |
| **sgTSPAN15_5** | GTCTGTGGGCATCTATGCAG |
| **sgTTI2_1** | GGGGCAGGTAGCAAAAGCCC |
| **sgTTI2_2** | GACCTCATCACAATGGGTGG |
| **sgTTI2_3** | AGGTCCTGCGTAAAAGAAGG |
| **sgTTI2_4** | TAAGAGCGCCCTGCTACAGG |
| **sgTTI2_5** | ACACTGTCTTGCCCGCCCGG |
| **sgTTL_1** | GAACAGCAGCGTCTACGCCG |
| **sgTTL_2** | GGGCTCACCCAGTCTCCCGA |
| **sgTTL_3** | GGCTTACTGAGCACATGCAG |
| **sgTTL_4** | CAGTTGGTGAATTACTACAG |
| **sgTTL_5** | GCTGGAAGCTCTGGTAAGGG |
| **sgUSP45_1** | AAGAAACAGGAACAAGGCTG |
| **sgUSP45_2** | AGTGAGGACAGAAGAAACAA |
| **sgUSP45_3** | ACAGATCATGATCGATACAG |
| **sgUSP45_4** | GAAGCACTTTCAGATTCTGA |
| **sgUSP45_5** | GGGCTGTTCAGATCCAGTAG |
| **sgUTP15_1** | GACACCAGAAGACCTCCAGA |
| **sgUTP15_2** | GTCTGGGACATGTTAAAAGG |
| **sgUTP15_3** | GCATAATTATATGGAGGCTG |
| **sgUTP15_4** | AAAAGTAGGACACTCTCCAC |
| **sgUTP15_5** | ATGGCCTTCAAACTGCCTGA |
| **sgVASP_1** | GCTGAGAGTGGTCGAAGCGG |
| **sgVASP_2** | ATGATGATGGCAACAAGCGA |
| **sgVASP_3** | GACGCGAAAGGAATTGGCCG |
| **sgVASP_4** | GATGCGGCCCAGTTTGCCGC |
| **sgVASP_5** | CGTCGGAGCACATAGAGCGC |
| **sgVCP_1** | GCTGCCCATTGATGACACAG |
| **sgVCP_2** | GACACAGTGATCCACTGCGA |
| **sgVCP_3** | TGATAGGCTCCCCTTCGCAG |
| **sgVCP_4** | GTTCTGGCAGACTCATGGCG |
| **sgVCP_5** | TCATGGATGGCCTAAAGCAG |
| **sgWDFY4_1** | GATTCATCTCACACACACAG |
| **sgWDFY4_2** | GAGAAGGTCAGACCTCCAGG |
| **sgWDFY4_3** | CCAGAGGAAAGACAAAGAAG |
| **sgWDFY4_4** | GGAACAGAGCCGAGGAGACG |

| | |
|---|---|
| **sgWDFY4_5** | GCTGCCTTCAGAGTCTCCAG |
| **sgWDR77_1** | GCAGCGCCCACTCAGGCTGG |
| **sgWDR77_2** | CAACTGCCGTTCCATGCAGG |
| **sgWDR77_3** | GTGGTACTGAGTTCATACCG |
| **sgWDR77_4** | GGGAGTGTACAGCTGAGCTC |
| **sgWDR77_5** | TAGTGACTCACCTGAATCGG |
| **sgWRN_1** | GTAAATTGGAAAACCCACGG |
| **sgWRN_2** | GAAATGGCTGCAGATCCTGA |
| **sgWRN_3** | GGCAACAAACAAGATACTGG |
| **sgWRN_4** | TGGAAACAACTGCACAGCAG |
| **sgWRN_5** | CAGTGAATTCTAAGAAGGGG |
| **sgYEATS2_1** | AAGGGAACCCGAGAATGAGG |
| **sgYEATS2_2** | GCTCCCATAAAACAGAACAG |
| **sgYEATS2_3** | GGAATGGTCAAAGGCAAAGA |
| **sgYEATS2_4** | GCTGGAAAGCTGAATTGCCA |
| **sgYEATS2_5** | TTGGGAGACCTGAGAAGCCG |
| **sgYRDC_1** | GCCAGCGTGGGGTTGAGCGA |
| **sgYRDC_2** | GTACAGCGTATCGGTGGGGA |
| **sgYRDC_3** | GTGTCGCCTTGGCTCAACTG |
| **sgYRDC_4** | TCTCCGGCGCGTCGGTGCAG |
| **sgYRDC_5** | CGGCCGAGGCATACGGCCAG |
| **sgZBED2_1** | GAAGAGGTGACATGATGAGG |
| **sgZBED2_2** | GAGGCGGGAAGACGAGGAAG |
| **sgZBED2_3** | GGACTTAGAGATGAAGGAGG |
| **sgZBED2_4** | GAGCTGGAGAAGAGTGGCCA |
| **sgZBED2_5** | GGCGCTGCCCAGCCTGACCA |
| **sgZBTB32_1** | GCTAAGGCCCCTTCAGGAGG |
| **sgZBTB32_2** | GGAATCCTGAGAGAGAACTG |
| **sgZBTB32_3** | GAACAGGTTTCTAGAACTGG |
| **sgZBTB32_4** | AGGAGGTCTGGCGGGAACAG |
| **sgZBTB32_5** | GGAGTTCCCCGCCCACAGCC |
| **sgZNF292_1** | GAGAGGTTGAGTTGCGGCGA |
| **sgZNF292_2** | GGCGGACGAAGAGGCCGAGC |
| **sgZNF292_3** | GTCGCGGAGCTGCAGCGCCT |
| **sgZNF292_4** | CAGCTGCTGACAGTAGTCGG |
| **sgZNF292_5** | AGGAGCTGGAGCTACAGCTG |
| **CXCL8_1** | CTAAGTTCTTTAGCACTCCT |
| **CXCL8_2** | AAATTTGGGGTGGAAAGGTT |
| **CXCL8_3** | TTGATAAATTTGGGGTGGAA |

| | |
|---|---|
| **CXCL8_4** | GTTCTTTGATAAATTTGGGG |
| **OIP5-AS1_1** | AGAAAAAAAATATCGGCCAG |
| **OIP5-AS1_2** | AAAAATATCGGCCAGAGGAG |
| **OIP5-AS1_3** | CTACACTAAAGCAAAGTGTT |
| **HACD3_1** | GGTCAGTTTGTAAACAGGCT |
| **HACD3_2** | CTCTGGGTCAGTTTGTAAAC |
| **HACD3_3** | CTGTTTACAAACTGACCCAG |
| **HACD3_4** | TTACAAACTGACCCAGAGGC |
| **LINC00599_1** | CCAGAAAGCTTTAATGTGTG |
| **LINC00599_2** | CAGAAAGCTTTAATGTGTGT |
| **LINC00599_3** | TCTATTAGGTCTGTGCGCAT |
| **LINC00599_4** | TAGGTCTGTGCGCATGGGTG |
| **JADE3_1** | TGAGGTCCTTCCGGAATACC |
| **JADE3_2** | ATCTTTCCAGGTATTCCGGA |
| **JADE3_3** | TGGCACTGATGAGGTCCTTC |
| **JADE3_4** | GAAGTTTCATGGCACTGATG |
| **CFAP57_1** | TGCAGGCTCAGAGAAGAGTC |
| **CFAP57_2** | GCAGGCTCAGAGAAGAGTCA |
| **CFAP57_3** | AGAGAAGAGTCAGGGCATGT |
| **CFAP57_4** | GTACCGCCGATTGGGACTGA |
| **ABALON_1** | TGGCCTCATTTCACCACAAG |
| **ABALON_2** | TAACCTCTTGTGGTGAAATG |
| **ABALON_3** | GTGGTGAAATGAGGCCAGTC |
| **ABALON_4** | GATAGTTGAGGAAACCTGAC |
| **MLK7-AS1(MAP3K20-AS1)_1** | CTCAGCCATCAGTGGCAACT |
| **MLK7-AS1(MAP3K20-AS1)_2** | GCCATCAGTGGCAACTTGGT |
| **MLK7-AS1(MAP3K20-AS1)_3** | CCATCAGTGGCAACTTGGTC |
| **MLK7-AS1(MAP3K20-AS1)_4** | TCGGGTCTGCTTCCGCGCTG |
| **TIGAR_1** | TCTTAATATATCACAGGACA |
| **TIGAR_2** | TTAAATCCAGTTTCTGAAAG |
| **TIGAR_3** | TTTAAACAAGCAGCAGCTGC |
| **TIGAR_4** | CTTTGTCCTCATGAGATCAC |
| **APOL4_1** | AGGCTTCATCGCTAGTCAGC |
| **APOL4_2** | TGCTGACTAGCGATGAAGCC |
| **APOL4_3** | CTGGAAGAGATTTGTGCGTG |
| **APOL4_4** | TGCGTGTGGCTGAATTGCCC |

| | |
|---|---|
| **TOX1_1** | GTCTTTTCACAGTCCTACCC |
| **TOX1_2** | CACTTTCCAGGCTTGGACCA |
| **TOX1_3** | TCACTTTCCAGGCTTGGACC |
| **TOX1_4** | GTCCTACCCTGGTCCAAGCC |
| **Lag-3_1** | ACCTCAGCCCCTGGCTGGAG |
| **Lag-3_2** | CTAGTGAAGCCTCTCCAGCC |
| **Lag-3_3** | ACCACCGGGACCTCAGCCCC |
| **Lag-3_4** | GGGCTGAGGTCCCGGTGGTG |
| **Tim-3_1** | CGCTCTGTATTCCACTTCTG |
| **Tim-3_2** | CTATGCAGGGTCCTCAGAAG |
| **Tim-3_3** | CTCAGAAGTGGAATACAGAG |
| **Tim-3_4** | AGAAGTGGAATACAGAGCGG |
| **Ikzf2_1** | ACAATGAGCTTTCACCCGAA |
| **Ikzf2_2** | CAATGAGCTTTCACCCGAAA |
| **Ikzf2_3** | CGAAAGGGAGCACTCCAATA |
| **Ikzf2_4** | GGTGAGGTCAATTGCCATAT |
| **Setbp1_1** | TAGAGATTGTCATGGAGTCC |
| **Setbp1_2** | AGAGATTGTCATGGAGTCCA |
| **Setbp1_3** | TGACTCGCCCCCTCTTTGCC |
| **Setbp1_4** | TTAAGCAGCTCCCGGCAAAG |
| **NFATc1_1** | TTTTTCTCTCTAGAACACTA |
| **NFATc1_2** | CAGGGCGGGGCTGACGTTGG |
| **NFATc1_3** | CGGCAGGGCGGGGCTGACGT |
| **NFATc1_4** | GTGCGCCGTGGGGAGCGGCA |
| **Blimp-1(PRDM1)_1** | AGTTACACTTGGGGGCAGCC |
| **Blimp-1(PRDM1)_2** | AGTGCTGGAGTTACACTTGG |
| **Blimp-1(PRDM1)_3** | CAGTGCTGGAGTTACACTTG |
| **Blimp-1(PRDM1)_4** | ACAGTGCTGGAGTTACACTT |
| **Eomes_1** | AACACATTGTAGTGGGCAGT |
| **Eomes_2** | GAACACATTGTAGTGGGCAG |
| **Eomes_3** | CTCTACGAACACATTGTAGT |
| **Eomes_4** | CCACTACAATGTGTTCGTAG |
| **sgADAM8_1** | GGACCACAGCAAGAACCCCG |
| **sgADAM8_2** | GGGCTGGACACAGTGACCGG |
| **sgADAM8_3** | GAGGTGACGGAGCAGCCTCG |
| **sgADAM8_4** | AGGCACGGCAACGGACACGG |
| **sgADARB1_1** | CAGAAAGCGGCCCCTGGAGG |
| **sgADARB1_2** | AGTACCGCCTGAAGAAAAGG |
| **sgADARB1_3** | AGAAAAGGAGGAAAACACCA |

| | |
|---|---|
| **sgADARB1_4** | GGTGGTAAGACAGGGAGAGG |
| **sgALDOC_1** | CAAGCGGCTGAGCCAAATTG |
| **sgALDOC_2** | AATTGGGGTGGAAAACACAG |
| **sgALDOC_3** | AAAACACAGAGGAGAACCGC |
| **sgALDOC_4** | GGACCTGGCGGTACAGCCGG |
| **sgANKZF1_1** | GGAGATCGACGCAGGAGCCG |
| **sgANKZF1_2** | GCAGACACTGGATCGGGAGA |
| **sgANKZF1_3** | CAGGGGCTTCGGGATGCCCG |
| **sgANKZF1_4** | GCAGTCTGACTGCTTCCCGA |
| **sgANXA1_1** | GAGCTAAAACAACCTCCTCA |
| **sgANXA1_2** | ACTGTGAAGTCATCCAAAGG |
| **sgANXA1_3** | AAGTTGTGGATAGCTTCTGG |
| **sgANXA1_4** | GGATGGATTGAAGGTAGGAT |
| **sgAPOBR_1** | GGTAGGAGACAAAGGTGCCG |
| **sgAPOBR_2** | GGGGCTTGCCAAGACAGGAG |
| **sgAPOBR_3** | GGAGCGGCCAAGCCCAGGAG |
| **sgAPOBR_4** | GAGAAGCTGGGAACAGGAGG |
| **sgARHGAP4_1** | GCAGTGCAGCCCTCCACCAG |
| **sgARHGAP4_2** | GTGGAGCTGGAATACTCCCG |
| **sgARHGAP4_3** | GCAGTGCAAGGGCGACAGGA |
| **sgARHGAP4_4** | GCGCCTGAGTCACATTGCAG |
| **sgARHGAP9_1** | GGTGAGCGGGAACTTGGCAG |
| **sgARHGAP9_2** | GGGGTGGGGACCGAGGACAG |
| **sgARHGAP9_3** | GCTGCGAGGTGCGTTGGCTG |
| **sgARHGAP9_4** | AAGTAGCGTGGACCTGCGCG |
| **sgARHGEF39_1** | GCAGAGCCTGGACACGCCGA |
| **sgARHGEF39_2** | GGTCTCTAGCAGCTCCCGGG |
| **sgARHGEF39_3** | GGCGCTACCAAGAACAGCTG |
| **sgARHGEF39_4** | GCACCACTAACAGCCAGCCC |
| **sgBNIP3_1** | GAGAGAGCAGCAGAGATGGA |
| **sgBNIP3_2** | CATCACCTACCCCAATCCGA |
| **sgBNIP3_3** | GGATATGGGATTGGTCAAGT |
| **sgBNIP3_4** | GCTGAAGTGCAGTTCTACCC |
| **sgC1QTNF3_1** | GGAATGCCAGGAGGGCCCGG |
| **sgC1QTNF3_2** | TTTTAGTCTCCACAAACCGG |
| **sgC1QTNF3_3** | GCATGGCCCCAAAGGAGAGA |
| **sgC1QTNF3_4** | AAGGAGAGAAGGGCTACCCG |
| **sgC4orf3_1** | CTGCAGGAGCGAGATGGAGG |
| **sgC4orf3_2** | GGAGCGGCGAGGCTTTAGCG |

| | |
|---|---|
| **sgC4orf3_3** | CGAAAAGGATGAAAAGCCAG |
| **sgC4orf3_4** | GAAACACCACCACATCGAAA |
| **sgCACNA1I_1** | GAGGACGAGGCGGATGGGGA |
| **sgCACNA1I_2** | GAAGGCCTGGACAGCAGCGG |
| **sgCACNA1I_3** | GAGGTGTCGGTACCTGCCGG |
| **sgCACNA1I_4** | GGGGCTGCTGCAGATGAGGA |
| **sgCALCOCO1_1** | AAGTTGACTCCACCACGGGA |
| **sgCALCOCO1_2** | TGAACTGGTGACCCTGGAGG |
| **sgCALCOCO1_3** | GATGCAGCTGAAGCTACAGC |
| **sgCALCOCO1_4** | GCAGGAGCACACGGAGCTGA |
| **sgCAPN3_1** | GTGTTTGTCTGCGTACCGCA |
| **sgCAPN3_2** | GCATTTGACAAGGATGGAGA |
| **sgCAPN3_3** | CATCAAGCTCAACGTTCTGG |
| **sgCAPN3_4** | ACAAGGACCTGAAGACACAC |
| **sgCCDC112_1** | AAATTCCTTCAGCAAACAGG |
| **sgCCDC112_2** | TGAGAAGCTCAGAGAAATGA |
| **sgCCDC112_3** | AGGCTTTACAAACCCACCAA |
| **sgCCDC112_4** | CTGGTGATCATAATCATCCC |
| **sgCERCAM_1** | CAGGATGGCAAGGACCACGG |
| **sgCERCAM_2** | GGGCCGAGTCACATGAGCTG |
| **sgCERCAM_3** | CAGAGGGCCATCCTGGCCCG |
| **sgCERCAM_4** | GGAAGTACTCGGCTGTGCGG |
| **sgCHD3_1** | AGGAGAAGAAGACAAAGCGG |
| **sgCHD3_2** | CCAGTGGCTGTGATAAACCG |
| **sgCHD3_3** | GGAGCGCTTAGAATCGGCGC |
| **sgCHD3_4** | GCTGTAATTGGCGCCTGCGG |
| **sgCHN2_1** | GTGGATCATTAGGTACCGGA |
| **sgCHN2_2** | GGCAGCCCTTCGAACCAGGG |
| **sgCHN2_3** | CACAATATTCACACCAGTGT |
| **sgCHN2_4** | TGACCTCACAACACTTGTGA |
| **sgCLEC2B_1** | GGAAACCAATCCAATCATAG |
| **sgCLEC2B_2** | TCACTGGATTGGACTGAAGA |
| **sgCLEC2B_3** | AGTTAGGTCGGCATGTTGAG |
| **sgCLEC2B_4** | AGAGTTTATGCCCCTATGAT |
| **sgCTDSP1_1** | GGTTGGGTCGAGACCTGCGG |
| **sgCTDSP1_2** | AGAGTGAGTGGAGGATGCCC |
| **sgCTDSP1_3** | CCATTCTCCTCCACAAGCAG |
| **sgCTDSP1_4** | TCCCAGAAGCCCCGAAGCCG |
| **sgCX3CR1_1** | AAAGCCGATGAAGAAGAAGG |

| | |
|---|---|
| **sgCX3CR1_2** | GCACGGGCCAGATTTCCTGG |
| **sgCX3CR1_3** | CAACCGTCTCAGTCACACTG |
| **sgCX3CR1_4** | CCAGGCATTTCCCATACAGG |
| **sgDOK2_1** | GGAGCGCCTGTACCTCCTGG |
| **sgDOK2_2** | GGAGACGGGGCAGTGAAACA |
| **sgDOK2_3** | GCAGAGCCGGCCCTGCATGG |
| **sgDOK2_4** | AAAGCGCCGCAGAAACCTGT |
| **sgDTHD1_1** | GGATTCAGAAGCCAAGACAG |
| **sgDTHD1_2** | GGAACTCCAAATCAAAGAAG |
| **sgDTHD1_3** | GCAGCAGACATTGCTGCAAG |
| **sgDTHD1_4** | TGTTTCTCCCACAAATGGAG |
| **sgEFNA3_1** | TGGGCCAGCAGCGGCAGCAG |
| **sgEFNA3_2** | CCAGCCAGGGCTTCAAGCGC |
| **sgEFNA3_3** | GAGAGAAGGCGCTGTAGCGC |
| **sgEFNA3_4** | GTGCAGGTTGTGAGTGGGCG |
| **sgEFNA5_1** | GGACGGAGTCCTCATAGTGA |
| **sgEFNA5_2** | GCACGTGGAGATGTTGACGC |
| **sgEFNA5_3** | AGTGAGGGCAGAAAACATCC |
| **sgEFNA5_4** | CATCAAAGTTCACCATGTAG |
| **sgEGLN1_1** | GCCATGGCCAATGACAGCGG |
| **sgEGLN1_2** | CCGGGAGCCCAGGAAGGCAG |
| **sgEGLN1_3** | CAAGCACGGCATCTGTGTGG |
| **sgEGLN1_4** | GACTCGTCCAAGGACATCCG |
| **sgERP29_1** | GCGGGGAGAGAAATGCGGCG |
| **sgERP29_2** | CAGGAGAAGGGGAAGCAGCG |
| **sgERP29_3** | CGGCAGCGGCCTGCACACCA |
| **sgERP29_4** | TCCGCTCCGCATGGCGGCAG |
| **sgFAM117B_1** | GGCGGCGGCAACAACAACGG |
| **sgFAM117B_2** | GACACGGTGCCCAGCAGCGG |
| **sgFAM117B_3** | AGTGCGTCGGATAATACTGG |
| **sgFAM117B_4** | GAGCCGCTTGCAGCCCATGA |
| **sgFMR1_1** | GAAGATGGAGGAGCTGGTGG |
| **sgFMR1_2** | GGACAAGGAGGAAGAGGACG |
| **sgFMR1_3** | TTCTAGGTGTGCCAAAGAGG |
| **sgFMR1_4** | GCAAGCTAGAAAAGTACCTG |
| **sgFOXO4_1** | GCTGCTGCTATCCATGGAGG |
| **sgFOXO4_2** | GGTCATGAGGACGTCAGCAG |
| **sgFOXO4_3** | GGTGCTATCATAGAAAGGGT |
| **sgFOXO4_4** | GCACTTGCCATGACTGGAGG |

| | |
|---|---|
| **sgFOXP3_1** | GGGGTTCAAGGAAGAAGAGG |
| **sgFOXP3_2** | CAGGAGCCTCGCCCAGCTGG |
| **sgFOXP3_3** | GGGGTGCACCTGCAGCACAG |
| **sgFOXP3_4** | GAAGACCCCAGTGGCGGTGG |
| **sgGCHFR_1** | GCTGATGAGCAGGTAGGGCA |
| **sgGCHFR_2** | GGAGGTGGGCCCCACTATGG |
| **sgGCHFR_3** | GGCTTCAAAGAGAAGAGCCT |
| **sgGCHFR_4** | CATCAGCACCCAGATCCGCA |
| **sgGLB1L_1** | CAGTGGTATAGAGTCCCCGG |
| **sgGLB1L_2** | GCGGCAGCAGCAGGGAACGA |
| **sgGLB1L_3** | CGGTCGGCCCAAAGCACCCG |
| **sgGLB1L_4** | GCAGGCTGCCAGACACATAG |
| **sgGPI_1** | TGAGAAGATCAACTACACCG |
| **sgGPI_2** | GATGAGGGGAAAATCGACGG |
| **sgGPI_3** | GGAGGCAATGATGAACAGGG |
| **sgGPI_4** | AAACACACTCACCCACGTGC |
| **sgGPR137C_1** | GGAGCCGCTGAGCGAGAAGG |
| **sgGPR137C_2** | GAAGGCGGCGGAGAAGAGGG |
| **sgGPR137C_3** | GGAAGCAGTAGAGCAGCCAG |
| **sgGPR137C_4** | GCGCTGCCCACAGGAGACAG |
| **sgGPR155_1** | GCAGAACAGGGTGACTGCTG |
| **sgGPR155_2** | GTCTCAGAGCCACAAAGTGG |
| **sgGPR155_3** | TTTGCAGATAACCTCAGGGA |
| **sgGPR155_4** | CCAGAGCTGCATATTAGCCC |
| **sgIL16_1** | AGAGGGTCTCACGGAAGCCA |
| **sgIL16_2** | GATTCTCTGCCTGATGGAGG |
| **sgIL16_3** | GAGGGGACTCAGACACACCT |
| **sgIL16_4** | GTTGACAGCAGCCTTAGGAG |
| **sgIL18RAP_1** | AAGTGATGCACAAAGTCCAG |
| **sgIL18RAP_2** | CAACTGAAAGAAAAGAGAGG |
| **sgIL18RAP_3** | GAAAGAGATGTGGCTCCAGG |
| **sgIL18RAP_4** | GGATCCTGTCGAGGACACAC |
| **sgIL24_1** | GGGTTGTTCCCCAGAAACTG |
| **sgIL24_2** | TCAGAGACAGTGCACACAGG |
| **sgIL24_3** | GAGGCTGTCGCCAGCAAAGG |
| **sgIL24_4** | AGCAGGGTAAAACCCAGGCA |
| **sgINSIG2_1** | AAGGAGAGACAGAGTCACCT |
| **sgINSIG2_2** | CTAGATGTCTGTCAATGCAG |
| **sgINSIG2_3** | GAATCATCAAGTTCACACTC |

| | |
|---|---|
| **sgINSIG2_4** | AAAAGTCCACCACAGTCCAA |
| **sgIPCEF1_1** | GCCATGCTGACTGCCAAGGG |
| **sgIPCEF1_2** | AAAGTTCTGGGTGATACTGA |
| **sgIPCEF1_3** | GGAAGATCCAGAAATAGCTG |
| **sgIPCEF1_4** | GAGTCTGGGAAGCGTGAGGA |
| **sgIRAK4_1** | GGTGGTAATAAAATGGGAGA |
| **sgIRAK4_2** | CTCATGTGCCAAGAAAGTGG |
| **sgIRAK4_3** | AATAACTGGACTTCCAGCTG |
| **sgIRAK4_4** | GAACGACCCATTTCTGTTGG |
| **sgJAM3_1** | GCTCTGGAATCCGTGGGCAG |
| **sgJAM3_2** | CAAGTGACCCCAGGATCGAG |
| **sgJAM3_3** | CAGGTGTGAGGAGCAGGAGA |
| **sgJAM3_4** | CGAGCGCAGAGCCGGAGTCG |
| **sgJPH2_1** | AGGAGCCAGAGTATTCGCCC |
| **sgJPH2_2** | GGAGCCAGGGCAAACGGCAT |
| **sgJPH2_3** | CTGGAACTTTGGCTTTGAGG |
| **sgJPH2_4** | GGGCTGGGCATAGAGACCAA |
| **sgKDM4B_1** | GTATGATGACATCGACGACG |
| **sgKDM4B_2** | TCACCAGGTACTGTACCCCG |
| **sgKDM4B_3** | AGATCGGGGAGACAAAGGTG |
| **sgKDM4B_4** | GAAGCCGGCGTGGTAGCCGT |
| **sgKIAA0101_1** | TGTCGAATTAGTGGCAGAGG |
| **sgKIAA0101_2** | GTTCCAGGCACTTACAGAAA |
| **sgKIAA0101_3** | ATTCGACATCAGTTTCATCG |
| **sgKIAA0101_4** | AAGAACCAAGCACCTTTCTG |
| **sgKIF19_1** | AGGCAAAGAGAAGGAAGCGG |
| **sgKIF19_2** | GCTGCAATCATGAAGGACAG |
| **sgKIF19_3** | GCTCCAGGTAGCCCAGGGAG |
| **sgKIF19_4** | GGTGTTCCGGGACTCCTCGA |
| **sgLGALS8_1** | TGAAAGGCCACATCGGCTCG |
| **sgLGALS8_2** | GGAACTTTGATTGTGATACG |
| **sgLGALS8_3** | CAGCCGGCCCTTTTGAAACG |
| **sgLGALS8_4** | GATCGTGATTATGGTGCTGA |
| **sgLIME1_1** | GAGGCTGCAGGGCAGTGCGA |
| **sgLIME1_2** | AGACTGCACGAGCTGCACCG |
| **sgLIME1_3** | GTGGCTCTTGGGGACTGCGA |
| **sgLIME1_4** | GAGGCCACCTATTCCAACGT |
| **sgMAPK13_1** | GAAGATGCTGGAGCTAGACG |
| **sgMAPK13_2** | GGAAGGGTTCAAAGAAGGGA |

| | |
|---|---|
| **sgMAPK13_3** | GGACCCTGAGGAAGAGACGG |
| **sgMAPK13_4** | ATAGAAGTTGCGCAGGGAGG |
| **sgMATK_1** | AGCACACCCGCCCCAAGCCA |
| **sgMATK_2** | GGGGAGCTGGCCTTCCGCAA |
| **sgMATK_3** | GCCGAGGAGGAGCTGGCCAG |
| **sgMATK_4** | GAGCTGAGCGGTGTTCACGA |
| **sgMGAT1_1** | GATGGGAATCACCGCCGGCG |
| **sgMGAT1_2** | GATGGGGAAGAGCTCAGCCG |
| **sgMGAT1_3** | GCAGTCCTGGCTAACGATGA |
| **sgMGAT1_4** | AGTAGCCCTGGAACTTGCGG |
| **sgMKNK2_1** | GCCTCACCTTGAACGAACGG |
| **sgMKNK2_2** | TGAGCCACATCCACAAGCGC |
| **sgMKNK2_3** | CAGCGGCATCAAACTCAACG |
| **sgMKNK2_4** | GGCGGAGTACATGGCCCCGG |
| **sgMOB3A_1** | GACTTGCTTCAGGAAGGGGT |
| **sgMOB3A_2** | GCTCAAACTTGCGCTTGGGG |
| **sgMOB3A_3** | CGAGCTGCACAAGAAGGCGC |
| **sgMOB3A_4** | TGAACGCCGGGCTGGACCTG |
| **sgMXI1_1** | CGGGCACAGAAACACAGCAG |
| **sgMXI1_2** | GTGGCGACTGGAACAGCTGC |
| **sgMXI1_3** | GCTGCAGGGTCCTCAGGAGA |
| **sgMXI1_4** | TGCCCGGCTCAACCTCCGTG |
| **sgMYOM1_1** | GGAGCCCTGGGTGTAGACGG |
| **sgMYOM1_2** | GGAAGGCCTCGGACTCCCGG |
| **sgMYOM1_3** | GAGCACCACATAGCTCCGGG |
| **sgMYOM1_4** | GCACCGTGAGTCACTACCAG |
| **sgNMUR1_1** | GCAGAAGAAGAGCAGCGCGG |
| **sgNMUR1_2** | GAGCACGCTCATGATGGCCA |
| **sgNMUR1_3** | TCATTGGGCTGCGACTGCGG |
| **sgNMUR1_4** | GGCCTGGAGTGGGTGCACCA |
| **sgNSUN6_1** | GACTTGTGTGCAGCACCTGG |
| **sgNSUN6_2** | AAGGATTCTGGTAGAAATGG |
| **sgNSUN6_3** | GGTTAAACTTGATATGGTGG |
| **sgNSUN6_4** | GGAAGGCTTTATGAATAAGG |
| **sgNUGGC_1** | GGAGGAGCTGAGCAGAGAAG |
| **sgNUGGC_2** | GGAGCAGCCTCTGCAAGAAG |
| **sgNUGGC_3** | GCAGCGGTAAGAAGTCCTGG |
| **sgNUGGC_4** | ACAAGGAATGGCAGAAACGA |
| **sgP4HA1_1** | ACAGCGACAAAAGATCCAGA |

| | |
|---|---|
| **sgP4HA1_2** | GAAATGCTGTGCCGTGGGGA |
| **sgP4HA1_3** | TCAAAGACCTAGCAAAACCA |
| **sgP4HA1_4** | CAGATGCTTTCAAAGAGCTG |
| **sgPATL2_1** | AGAAGAAGAGAATGAAGGGG |
| **sgPATL2_2** | TCTGGACCCAGACCTAGAGG |
| **sgPATL2_3** | AAATGTAGGCCAGAAAGGGA |
| **sgPATL2_4** | GAGGAAGGGGAAGGCCCTGG |
| **sgPBXIP1_1** | GAAGCAGAAGGAACAGCTGG |
| **sgPBXIP1_2** | GGAGTGGAAAGGAAAAGTGG |
| **sgPBXIP1_3** | GGAGTCGGGGAGCAAGAAGG |
| **sgPBXIP1_4** | GCTGGGACCATGGATGGAGA |
| **sgPCMTD2_1** | TGGGTATCTCAGCTCCATGG |
| **sgPCMTD2_2** | TTTTGGTGTGAACCATGGGG |
| **sgPCMTD2_3** | AATCTGACCTTCTCTTCCAG |
| **sgPCMTD2_4** | CAATTCCCAGTAACAAAGGA |
| **sgPCOLCE_1** | GCAGGGCGCAGGCAGTGAGG |
| **sgPCOLCE_2** | GATGAGGGCACAGGAGGACG |
| **sgPCOLCE_3** | GTTCTCACCAGTGCCCGAGG |
| **sgPCOLCE_4** | GGACTCGGGCCAGTTGGGCG |
| **sgPDE3B_1** | TGGAAGAAGAAGCCGCGCGG |
| **sgPDE3B_2** | GGCGAGGACAAAGGCAGCCA |
| **sgPDE3B_3** | GGTGAGGAAGAAGAAGGCAC |
| **sgPDE3B_4** | GCAGCACCAGAACGCAGTGC |
| **sgPDK1_1** | GAAGCTGCGGCTGAAGCCGG |
| **sgPDK1_2** | TAGATGAGTGACCGAGGAGG |
| **sgPDK1_3** | GCGCGGGAGGTCTCAACACG |
| **sgPDK1_4** | GACTCGGCATGAGGCTGGCG |
| **sgPDP1_1** | GCAGGCCAAAGGAGAACTGG |
| **sgPDP1_2** | GGTGGCAGTACACCCAAGGA |
| **sgPDP1_3** | GCAACCTGCTTGCAGACCAG |
| **sgPDP1_4** | GCTAGAGATTGAAAATGCAG |
| **sgPEG10_1** | GGAGCGCTCCAGCTTTGCTG |
| **sgPEG10_2** | GGTAGATCCAACCGAGCCGG |
| **sgPEG10_3** | GAAGGTCATGAAGCAGTCGG |
| **sgPEG10_4** | CCAGGTGCAGAAGCTCACAG |
| **sgPIK3CG_1** | GCTGGAGACCAGCGTGGCGG |
| **sgPIK3CG_2** | GAAGCTCTGCAGGATGGCGC |
| **sgPIK3CG_3** | GGAGAACTATAAACAGCCCG |
| **sgPIK3CG_4** | TGAGAGAGGACAACTGCCGA |

| | |
|---|---|
| **sgPLIN2_1** | GCAGAGTGTGGTGACTCGGG |
| **sgPLIN2_2** | AGCACTGGTCATGGCCACGG |
| **sgPLIN2_3** | ATGATGCAGCTCGTGAGCAG |
| **sgPLIN2_4** | GTAGAGTGGAAAAGGAGCAT |
| **sgPOLD4_1** | GGCCCGCTGGGCACAGCAAG |
| **sgPOLD4_2** | GATGAGCCGCTTCCGGCCCA |
| **sgPOLD4_3** | CCGGTTGTGAAGAGGAGGGA |
| **sgPOLD4_4** | GCTCCGCTTCCTCCTCGTCG |
| **sgPPP2R5B_1** | GGGCTCTGCGGAGGGAACGG |
| **sgPPP2R5B_2** | GAGGAGCTGTGGGAGCGGCG |
| **sgPPP2R5B_3** | GCCAGCTCACCTTTGAGCAG |
| **sgPPP2R5B_4** | TGTGGCCGACCTCAAGGGGA |
| **sgPRKACB_1** | TCAGGGCTACAATAAGGCAG |
| **sgPRKACB_2** | AGATGAGGTCTAGTGAATGG |
| **sgPRKACB_3** | GAGCTGCATAGAACCGTGCA |
| **sgPRKACB_4** | AAAATGGGAGAATCCAACTC |
| **sgPRR5L_1** | GCAGAATGGGAGCGAAGCCG |
| **sgPRR5L_2** | GGAGAGTCTCAGTGAAGAAG |
| **sgPRR5L_3** | GAGCTGCTTGGAATCGGGGA |
| **sgPRR5L_4** | GAGAAGATCAAGCTGTGTGA |
| **sgRAB24_1** | GCAGCGCGTGGACGTCAAGG |
| **sgRAB24_2** | GATAAGGCCCCACCAGAAAG |
| **sgRAB24_3** | GGAACTGCGCAGCCTAGAGG |
| **sgRAB24_4** | CGTGGGCAAGACTAGCCTGG |
| **sgRARRES3_1** | GGAGCGCCTGGAAGATGTGG |
| **sgRARRES3_2** | GATTGCAGGTGAGTACCCCG |
| **sgRARRES3_3** | CAGGACTGAGAAGACACTGG |
| **sgRARRES3_4** | GTACCAACCACGGCCCGTGG |
| **sgRASSF1_1** | GGAGGGTGGCTTCTTGCTGG |
| **sgRASSF1_2** | GCTTGACAGCATCCTTGGGC |
| **sgRASSF1_3** | TGGATGATGAGCAGCCCCTG |
| **sgRASSF1_4** | GAAGGAAAATGACTCTGGGG |
| **sgS100A10_1** | GGAGGACCTGAGAGTACTCA |
| **sgS100A10_2** | ATCTCAAATGGAACACGCCA |
| **sgS100A10_3** | CTCATGGAAAAGGAGTTCCC |
| **sgS100A10_4** | GACCTGGACCAGTGTAGAGA |
| **sgSAMD3_1** | CAAGAAGCAAAGCATTACAG |
| **sgSAMD3_2** | ATCCTCATCCAGGAAAGGGT |
| **sgSAMD3_3** | TGTAAATTTGGACACCGAAG |

| | |
|---|---|
| **sgSAMD3_4** | CCCATGCCTGGCTGGACTGG |
| **sgSAP30_1** | CCTGACGAGATGAGCCGCGG |
| **sgSAP30_2** | GCAGGCAGCACAGTTGCCCG |
| **sgSAP30_3** | GAAGAAGGTGAAGATCGAGC |
| **sgSAP30_4** | ACCGCTCACCATCCTCCCGC |
| **sgSASH3_1** | AAAAGCTGGGGAAGAAGTGG |
| **sgSASH3_2** | GATGGGCAAGATGATGGTGA |
| **sgSASH3_3** | CCTGCAGGCAGACACTCTGG |
| **sgSASH3_4** | GTAGTCTGGAGATGTCGGGG |
| **sgSCML4_1** | ACAGCTTGGCGAGGAAGCGG |
| **sgSCML4_2** | AGTGCGGGGACCACCAGCGG |
| **sgSCML4_3** | AGGATCCCGGGGAGAAAGCG |
| **sgSCML4_4** | GCGTGGAGTGAAGTGAGGGT |
| **sgSFR1_1** | GGGAGATGATGGATTCGCAG |
| **sgSFR1_2** | AGAAAGAAGACCTTCTTCGG |
| **sgSFR1_3** | GGTTTTACCTAGCACTCCTC |
| **sgSFR1_4** | CTAGTTTTAGCCTCCGAAGA |
| **sgSH3D21_1** | GGAGAGTGTCAAGGACGAAG |
| **sgSH3D21_2** | GCAGAGGCCTCTGAGAGAGG |
| **sgSH3D21_3** | GGAGGAGCTGAAGAGCGAGA |
| **sgSH3D21_4** | CAGACGAGTTGGCGCTGCGG |
| **sgSLC14A1_1** | ATCCCTTCCAGACAAACCCG |
| **sgSLC14A1_2** | AAGGGAGGGTGAAGACGGGG |
| **sgSLC14A1_3** | GCTGTCCACTCTAACCATAG |
| **sgSLC14A1_4** | ACACCACTTGGGATATGCCC |
| **sgSLC16A3_1** | GGTGCTGCTGATGGAGGCGG |
| **sgSLC16A3_2** | GGAGGGGCCGTGGTGGACGA |
| **sgSLC16A3_3** | GAGCATGGCCAGCAGGATGG |
| **sgSLC16A3_4** | CAGGTCTAGCAGGCGCCGGG |
| **sgSLC2A13_1** | AGTGCCAGTGTACATTGCGG |
| **sgSLC2A13_2** | AGAATGTGGAGTACACGCTG |
| **sgSLC2A13_3** | GAAGCAGCCGGAGCCGGACG |
| **sgSLC2A13_4** | GTTTGGCTATGACACCGGGG |
| **sgSTK10_1** | GGAGCTTCACAATGTAGGGG |
| **sgSTK10_2** | GCTGCAGGTGAAGAACGAGG |
| **sgSTK10_3** | AAGAGCCTCCACATCAACGG |
| **sgSTK10_4** | GCGGAAATTGGCAAAAGCCA |
| **sgSYF2_1** | GGTGCTGGTGGACAGCGCGG |
| **sgSYF2_2** | GGAGGAGGGGTCCCTCGCTG |

| | |
|---|---|
| **sgSYF2_3** | GTGCAGAAGATGCAGAAAGA |
| **sgSYF2_4** | GACCTGAAAATCCCAGATCA |
| **sgTC2N_1** | CAGGGGAGTAACAGAAGCCT |
| **sgTC2N_2** | GCATTTCAAATCTTCAGCCA |
| **sgTC2N_3** | AACATAGATTTGCTTAGCCC |
| **sgTC2N_4** | AGGAGTGTCTCCATAACTAG |
| **sgTFEC_1** | CATTCTACAGTGATATGCGC |
| **sgTFEC_2** | CAGACAGAAGAAATTAGAGC |
| **sgTFEC_3** | CATCAGTGGAGTACATCAAG |
| **sgTFEC_4** | GATGTGTATAGCGGTGAACA |
| **sgTMEM107_1** | GACGCTCCTGGCGCATCTGG |
| **sgTMEM107_2** | CAGGAGCGTCAGGAAGCGAG |
| **sgTMEM107_3** | GCCCAGGGTGACAGAGAGCG |
| **sgTMEM107_4** | CACCCTGGGCCTCTTTGCAG |
| **sgTMEM200B_1** | GAGCCCCGAAGAATGCGGGG |
| **sgTMEM200B_2** | GAGGTGCGGAGGAGCCCCGA |
| **sgTMEM200B_3** | GGTGGGTATGGGCATTGCAG |
| **sgTMEM200B_4** | GAGCTGCTGGGGCCCCAAGG |
| **sgTMEM45A_1** | GGGGAGCAGTTTATTCCTGG |
| **sgTMEM45A_2** | GGAACTCTAGGAAGGCAACG |
| **sgTMEM45A_3** | ACTGGAATCAACTCCTGGGC |
| **sgTMEM45A_4** | GGACACAGGAAGTGAACTGA |
| **sgTMEM71_1** | GTTCTCCTTATACATAACGC |
| **sgTMEM71_2** | GTATAGTGGGAGCCTGTCAG |
| **sgTMEM71_3** | GTGAAGACAACTGGTTGAAG |
| **sgTMEM71_4** | ACTCACCATTTCCATTGCAA |
| **sgTMEM9_1** | TGAGCAACTGCACAATGAGG |
| **sgTMEM9_2** | TAGCAACTGCCTGCACGTGG |
| **sgTMEM9_3** | AGTGCCTGGCCATGACGTGG |
| **sgTMEM9_4** | GAGCAACAGGGCACCCACCA |
| **sgTRADD_1** | GGAGTCCTCGCTGGACAAGG |
| **sgTRADD_2** | GCGCGCAGCTCCAGTTGCAG |
| **sgTRADD_3** | AAAGTCTGGGCAGGTGGCGG |
| **sgTRADD_4** | AATGGGCACGAAGAGTGGGT |
| **sgUBE2E1_1** | GAAGTGAGCAGATAGAAAGG |
| **sgUBE2E1_2** | CCAGGATCCGTGTATGAGGG |
| **sgUBE2E1_3** | GAATACACCACCCTCATACA |
| **sgUBE2E1_4** | GGGAGATTTCTTACCTTTGG |
| **sgWNT10B_1** | ACGGTGTTGGCCGTCAGCGG |

| | |
|---|---|
| **sgWNT10B_2** | GATGTGCAGACCCTGAAGCG |
| **sgWNT10B_3** | GCGAATCCACAACAACAGGG |
| **sgWNT10B_4** | GCAGCTGGTGCTGACACTCG |
| **sgWSB1_1** | CTGGAGGAGCAAATGACCGG |
| **sgWSB1_2** | AAGATACAGATCGTACCCAC |
| **sgWSB1_3** | GAGGTTCTGGAGAATTGATG |
| **sgWSB1_4** | AACACTGCCATCAGTAGAGA |
| **sgYPEL1_1** | GCTGCGGCCCTGCAGAGGAG |
| **sgYPEL1_2** | ACTGCAAGACCACGCTCGGG |
| **sgYPEL1_3** | GTACTCACACGGAATTGAAG |
| **sgYPEL1_4** | TGGCCAGGTGTGCTCTGCAG |
| **sgZBP1_1** | GCATGGACGATTTACCGCCC |
| **sgZBP1_2** | AAGACAGACAGTCTCCAGGG |
| **sgZBP1_3** | GCCCGGGTCAGCAGGAGCCT |
| **sgZBP1_4** | GTTGGCTGAACTGAGGGCCA |
| **sgZBTB1_1** | GTGTGGACTTACAATAACCG |
| **sgZBTB1_2** | ATGGCAATGAAGCCAACAGG |
| **sgZBTB1_3** | GCACAGACGGACAAATACAG |
| **sgZBTB1_4** | GTAAGTGCTGATGAAAGAGG |
| **sgZFP36_1** | GTAGATGGCAGTCAGATCCA |
| **sgZFP36_2** | GCGGGGACTCACCTCGTAGA |
| **sgZFP36_3** | CAGATCCATGGTGTAACGGT |
| **sgZFP36_4** | GTGGTGGTGGTGAGGTCCGG |
| **sgZNF395_1** | GGAGCGGCCCCTTCTAGCAG |
| **sgZNF395_2** | TCAGAGGTGGTGAAAGGCTG |
| **sgZNF395_3** | GGAGCAGCACAGCTGGATGG |
| **sgZNF395_4** | GCTGCAGGTCTGCTGCAGGG |
| **sgZSCAN25_1** | GGAATACCCAACTGCTGCTG |
| **sgZSCAN25_2** | GCCATGGGGCAGAGGAAGGG |
| **sgZSCAN25_3** | GGACCTGACAGAAAGAGCAC |
| **sgZSCAN25_4** | AAGAGCACTGGAGGCCAAGG |