# The regulatory role of *Malat1* on the alternative splicing factor SRSF1 during CD4+ T cell differentiation.

**Jack Greaves**

MSc by Thesis

University of York

Hull York Medical School

## Abstract

The proper activation and subsequent differentiation of naïve CD4+ T cells into effector T helper and regulatory T cells is vital for directing an appropriate adaptive immune response to specific infections. Recent evidence has identified long non-coding RNAs as novel regulators of CD4+ T cell activation and differentiation. Work by the Lagos group and others has identified the long non-coding RNA *Metastasis associated lung adenocarcinoma transcript 1 (Malat1)* as a critical regulator of Th cell function and immune response to chronic infection in mice. However, the mechanism behind this regulation by *Malat1* is not yet fully understood. This project aimed to investigate the RNA binding protein and splicing factor SRSF1, a known *Malat1* binding partner and prominent regulator of gene expression and alternative splicing in the immune system, as a mediator for *Malat1* regulation of Th cell function. Through analysis of individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP), we have shown that SRSF1 displays alternative RNA binding behaviour in Th2 cells upon *Malat1* loss. This alternative binding is directed towards RNA transcripts involved in T cell activation and differentiation, including *Il2ra* and *Runx3*. Following this, we found that *Runx3* abundance is reduced and isoform usage is altered upon *Malat1* loss in Th2 cells. To complement studies within *Malat1*$^{-/-}$ CD4+ T cells we attempted to develop *Srsf1*$^{-/-}$ CD4+ T cell models. Attempts to develop Srsf1$^{-/-}$ EL4 cell lines using CRISPR-Cas9-editing caused just a transient knockdown of SRSF1 expression, suggesting SRSF1 is essential for viability in this mouse T cell lymphoma cell line. Initial studies for establishing CRISPR-Cas9 RNP transfection into *in vitro* primary CD4+ T cell activation assays were unsuccessful at producing an SRSF1 knockout but have laid a foundation for further optimisation. Overall, our results identified *Malat1* regulation of SRSF1 mediated RNA interaction during the Th2 cell differentiation, which serves as a promising mechanism for further investigation to better characterise *Malat1* regulation of Th cell phenotype and cytokine expression.

# Table of Contents


Abstract ..... 2

List of Figures ..... 5

List of Tables ..... 7

Acknowledgements ..... 7

Declaration ..... 7

**1. Introduction ..... 8**

1.1 CD4+ T cell activation and differentiation ..... 8

1.2 Non-coding RNA ..... 10

1.2.1 Small non-coding RNAs ..... 11

1.2.2 Long non-coding RNAs ..... 11

1.2.3 LncRNAs are regulators of CD4+ T cells ..... 12

1.3 *Malat1* is a critical regulator within cellular stress and disease states ..... 13

1.3.1 *Malat1* is a driver of metastasis in a range of cancer types ..... 14

1.3.2 *Malat1* is a regulator of Th cell phenotype and host immunity towards infection. ..... 14

1.4 *Malat1* mechanisms of action ..... 16

1.4.1 Transcriptional, post-transcriptional and translational regulatory mechanisms of *Malat1*. ..... 16

1.4.2 *Malat1* regulates organisation of nuclear speckles through riboregulation of SR proteins 17

1.5 The SR protein family ..... 18

1.5.1 SRSF1 mechanisms of action ..... 19

1.5.2 SRSF1 is a negative regulator of CD4+ T cell activation and differentiation ..... 19

1.6 My Project ..... 20

1.7 Project Hypothesis and Aims ..... 21

**2. Materials and Methods ..... 21**

2.1 iCLIP ..... 21

2.2 Cell Culture ..... 22

2.3 I*n vitro* CD4+ T cell activation ..... 22

2.3.1 Sample collection ..... 22
2.3.2 CD4+ T cell isolation ..... 23
2.3.3 Naïve CD4+ T cell activation with Th2 polarisation conditions ..... 23
2.4 CRISPR-Cas9 genome editing ..... 24
2.4.1 CRISPR-Cas9 RNP production ..... 24
2.4.2 EL4 cell line electroporation mediated CRISPR-Cas9 RNP transfection ..... 25
2.4.3 Naïve CD4+ T cell electroporation mediated CRISPR-Cas9 RNP transfection ..... 25
2.5 Gapmer mediated SRSF1 knockdown in EL4 cells. ..... 26
2.6 RNA quantification ..... 27
2.6.1 RNA extraction ..... 27
2.6.2 Reverse transcription/ cDNA synthesis ..... 27
2.6.2 qPCR Primer design and validation ..... 28
2.6.4 Quantitative Real Time polymerase chain reaction ..... 28
2.7 Western Blotting ..... 28
2.7.1 Sample collection and protein extraction ..... 28
2.7.2 Protein sample concentration quantification – BCA assay ..... 29
2.7.3 Gel electrophoresis and membrane transfer ..... 29
2.7.3 Antibody incubation and chemiluminescent quantification ..... 30
2.8 Flow Cytometry ..... 30
2.8.1 Cellular staining ..... 30
2.8.2 Fluorescence quantification ..... 31
2.9 Statistics ..... 32
**3. Results ..... 33**
3.1 Analysis of alternative binding behaviour of SRSF1 in *Malat1*$^{-/-}$ Th2 cells ..... 33
3.1.1 General characteristics of SRSF1 crosslink events within Th2 cell iCLIP dataset ..... 34
3.1.2 Distribution of SRSF1 binding across the transcriptome. ..... 34
3.1.3 SRSF1 binding correlates with endogenous transcript abundance ..... 37
3.1.4 Greatest SRSF1 crosslink signal transcripts are significantly enriched in T cell function ..... 38

3.1.5 *Malat1* regulates SRSF1 RNA binding behaviour in Th2 cells ..... 38

3.2 Assessment of transcriptional and post-transcriptional regulation of *Il2ra* and *Runx3* by *Malat1* in Th2 cells. ..... 40

3.2.1 *In vitro* Th2 differentiation of naïve CD4+ T cells ..... 41

3.2.2 *Malat1* loss inhibits SRSF1 binding to *Il2ra* but does not affect *Il2ra* abundance or transcript usage ..... 41

3.2.3 *Malat1* loss increases SRSF1 binding to *Runx3* whilst reducing *Runx3* abundance and altering isoform usage ..... 45

3.3 Production of *in vitro* CRISPR-Cas9 mediated SRSF1 knockout models ..... 49

3.3.1 Optimisation of Electroporation conditions ..... 50

3.3.2 SRSF1 knockout in EL4 cells ..... 51

3.3.3 SRSF1 knockdown in EL4 cells ..... 53

3.3.4 SRSF1 knockout in naïve CD4+ T cells ..... 55

**4. Discussion ..... 58**

4.1 *Malat1* regulates SRSF1 binding to regulators of Th2 cell differentiation ..... 59

4.2 *Malat1* does not regulate *Il2ra* abundance through altered usage of the alternatively spliced *Il2ra* truncated isoform ..... 62

4.3 *Malat1* regulates *Runx3* expression in late stages of Th2 activation ..... 64

4.4 Electroporation mediated transfection of SRSF1 targeting crRNA associated RNPs reduces SRSF1 abundance in EL4 cells. ..... 67

4.5 Electroporation of naïve CD4+ T cells with CRISPR-Cas9 RNPs does not affect subsequent Th2 polarisation ..... 69

**5.0 Conclusion ..... 71**

**References ..... 72**

Appendix ..... 83

Appendix 1 ..... 83

## List of Figures

Figure 1. Schematic of CD4+ T cell activation and subsequent differentiation. ..... 10

Figure 2. Detection of SRSF1 enriched RNA transcripts in Th2 cells by iCLIP ..... 33
Figure 3. Two iCLIP replicates show high correlation of SRSF1 crosslink signal in WT and Malat1$^{-/-}$ Th2 cells..... 35
Figure 4. iCLIP data representing the RNA binding of SRSF1 shows a majority of transcripts with relatively low SRSF1 crosslink signal..... 36
Figure 5. Significant positive correlation is shared between transcript expression and transcript SRSF1 crosslink signal ..... 37
Figure 6. Transcripts with the greatest SRSF1 crosslink signal in WT Th2 cells are significantly enriched in T cell related gene sets. ..... 38
Figure 7. Malat1 regulates SRSF1 binding to transcripts significantly enriched in T cell related gene sets. ..... 40
Figure 8. rIL2 treatment of CD4+ T cells causes a significant increase in IL4 and IL10 expressing cells. ..... 41
Figure 9. Malat1 regulates SRSF1 binding with Il2ra specifically at the truncated Il2ra splicing site... 42
Figure 10. RT-qPCR primers sets were designed to assess Il2ra abundance and isoform usage. ..... 43
Figure 11. Malat1 does not regulate Il2ra expression or alternative splicing in Th2 cells. ..... 44
Figure 12. Malat1 regulates SRSF1 binding to the distal isoform of Runx3. ..... 46
Figure 13. RT-qPCR primers sets were designed to assess Runx3 abundance and isoform usage..... 47
Figure 14. Malat1 loss leads to lower pRunx3 isoform usage during Th2 activation. ..... 48
Figure 15. Malat1 regulates total Runx3 and dRunx3 isoform expression following IL2 treatment in differentiated Th2 cells. ..... 49
Figure 16. Electroporation of EL4 cells with HPRT targeting crRNA associated RNPs leads to a reduction in HPRT abundance. ..... 50
Figure 17. Optimisation is required to sustain acceptable El4 viability shortly after electroporation. 52
Figure 18. Electroporation mediated transfection of SRSF1 targeting CRIPSR-Cas9 RNPs into EL4 cells reduces SRSF1 abundance..... 53
Figure 19. Cloning by limiting dilution did not produce a clonal SRSF1 negative population. ..... 54
Figure 20. Gymnotic transfection of SRSF1 targeting gapmer into EL4 cells causes a knockdown of SRSF1 protein. ..... 55
Figure 21. Electroporation of CD4+ T cell does not negatively affect expression of IL4 and IL10 following Th2 polarisation in vitro. ..... 57
Figure 22. Transfection of SRSF1 targeting CRISPR-Cas9 RNPs to reduce SRSF1 abundance in Th2 cells was unsuccessful. ..... 58

## List of Tables

Table 1. crRNA sequences used in CRISPR-Cas9 RNPs for electroporation mediated transfection ..... 24
Table 2. Primers used for RNA quantification of targets in RTpPCR ..... 28
Table 3. Primary Antibodies used for Western Blotting ..... 30
Table 4. Recipes used for reagents/buffers made in house. ..... 32

## Acknowledgements

I thank my supervisor Prof. Dimitris Lagos for his expertise, considerate insight and dedicated time across the length of the project. I am also extremely thankful to Magnus Gwynne for their fantastic guidance throughout any and all practical work; as well as the whole of Q2 for making the length of the project an enriching and enjoyable experience.

## Declaration

I confirm that this work is original and that if any passage(s) or diagram(s) have been copied from academic papers, books, the internet or any other sources these are clearly identified by the use of quotation marks and the reference(s) is fully cited. I certify that, other than where indicated, this is my own work and does not breach the regulations of HYMS, the University of Hull or the University of York regarding plagiarism or academic conduct in examinations. I have read the HYMS Code of Practice on Academic Misconduct, and state that this piece of work is my own and does not contain any unacknowledged work from any other sources.

# 1. Introduction

## 1.1 CD4+ T cell activation and differentiation

CD4+ T cells are vital populations within the immune system for directed and specific response to a diverse range of infections. Their proper activation, represented by subtype-specific cytokine expression, is vital for correct action and subsequent resolution of the adaptive immune response. This activation is grounded in a large transcriptional shift consisting of a plethora of altered gene expression and alternative splicing (AS) events (Ip, et al., 2007).

Under homeostatic conditions CD4+ T cells remain in a naïve and inactive state. Upon infection, host naïve CD4+ T cells are activated through three specific molecular signals, leading to their differentiation into effector T helper (Th) cells or regulatory T cells (Tregs) across the course of infection. T cell receptor (TcR) of CD4+ T cells binding to a complementary antigen peptide presented by an antigen presenting cell (APC) represents signal 1. Conduction of signal 1 requires a co-stimulatory signal (CD80/86) on the APC surface binding to the CD28 receptor on the CD4+ T cell surface (signal 2). These initial activation signals are necessary for all naive CD4+ T cell activation events. The third signal is a specific polarising cytokine released following pattern recognition of invading pathogens by the innate immune system. Polarising cytokines drive activated CD4+ T cells to differentiate into the activated T helper (Th) subtype best suited to fight the invading pathogen detected by the innate immune system or infected tissue (Luckheeram, et al., 2012).

There a several different effector and regulatory CD4+ cell subtypes that differentiate from activated naïve CD4+ T cells, each defined by their own cytokine expression profile. A balance of these opposing effector and regulatory CD4+ subtypes across infection is important to prevent immunopathology and autoimmunity (Jäger & Kuchroo, 2010).

Effector Th cell populations mediate immune action against pathogens. IFNγ producing Th1 cells are the main population that mediate immunity against intracellular bacteria and parasites, usually through the action of promoting innate immune phagocytic action at the site of infection (Spellberg & Edwards Jr, 2001). Th2 cells produce IL4, IL5 and IL13 hallmark cytokines which can recruit mast cells and eosinophils to the site of infection as well as cause class switching of B-cell antibodies to IgE. This modifies the immune response specifically

towards clearing extracellular parasites (Lloyd & Snelgrove, 2018). Type 2 mediated immune response additionally drives allergic and autoimmune responses, therefore Th2 cells characterise several immunopathological disorders (Ko, et al., 2022) (Kubo, 2017). Additional to Th1 and Th2 cells, further effector CD4+ subtypes have also been characterised. Th17 cells express IL17A to mediate immunity towards a range of infections (ie. extracellular bacteria) as well as contributing to autoimmune inflammation (McGeachy & Cua, 2008) (Langrish, et al., 2005).

Regulatory T cells (Tregs) are vital for resolution of immune responses following effective pathogen removal (Tai, et al., 2019). Tregs express the anti-inflammatory cytokine IL10 as a hallmark cytokine, which limits the immune response mediated by effector Th cells and therefore prevents excessive damage to the host (Hori, et al., 2003). IL10 is also expressed by effector Th cells alongside their hallmark effector cytokines. IL10 expression provides a negative feedback regulator for the inflammatory response of all Th cells, regulating their action outside of the onset of immunopathology (Saraiva & O'Garra, 2010).

Due to the diversity, as well as emerging evidence of plasticity, both within and between CD4+ subtypes (Cano-Gamez, et al., 2020), rapid and sophisticated control over large transcriptional and post-transcriptional networks in response to polarising/effector cytokines is a requisite feature of CD4+ T cells. Extensive research has already been carried out on the internal molecular mechanisms of CD4+ T cells across activation and subsequent differentiation. The signal transduction pathways and transcription factors driving these cellular processes, are well understood (Luckheeram, et al., 2012). For example, the master linage transcription factor driving differentiation of Th1 (T-BET) (Szabo, et al., 2000), Th2 (GATA3) (Zhu, et al., 2006), Th17 (RORγt) (Yang, et al., 2008) and Treg (FOXP3) (Tai, et al., 2019) cells have been characterised and shown to be essential for expression of their respective CD4+ T cell subtype effector/regulatory cytokines. All information detailing the process of CD4+ T cell activation and differentiation, including the polarising cytokines, key transcription factors and effector cytokines of each activated Th/Treg subtype is summarised in Figure 1.

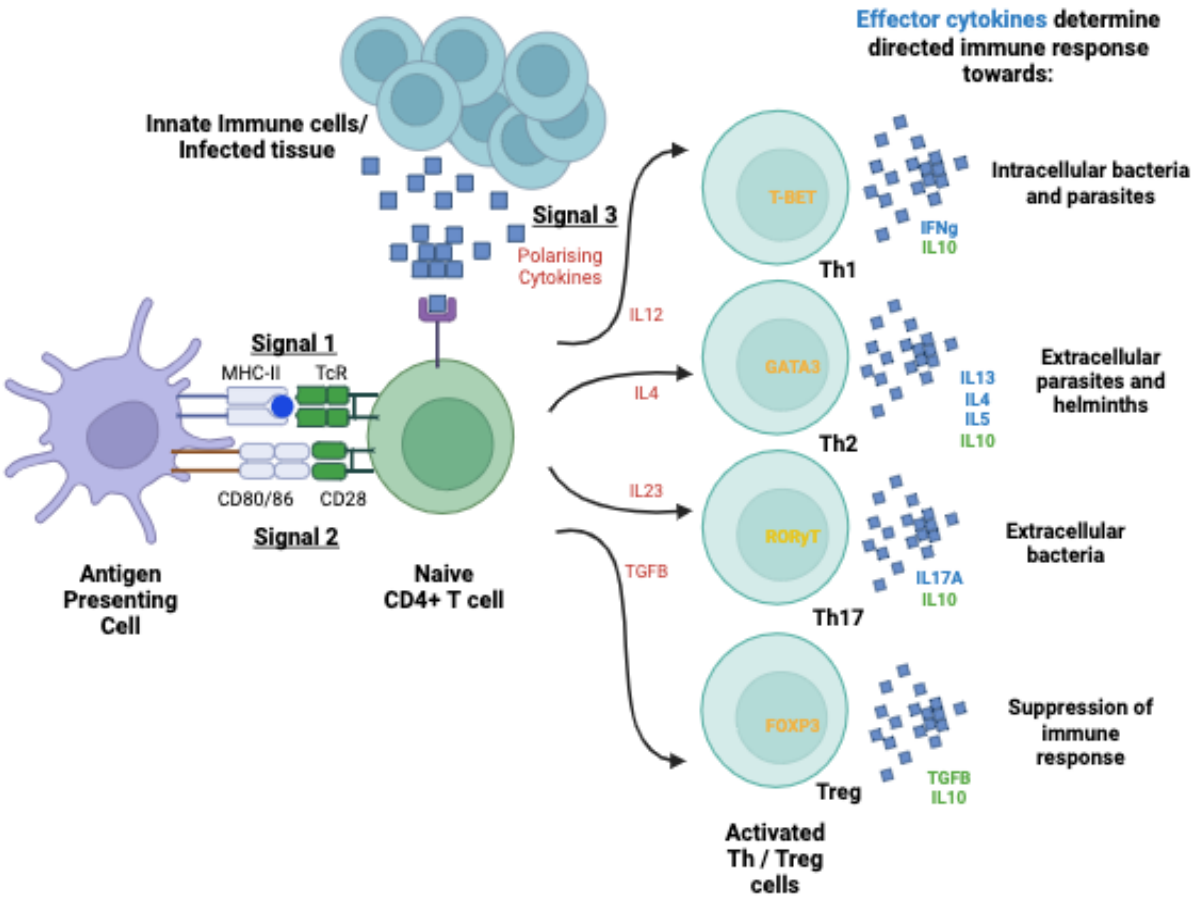


*Figure 1. Schematic of CD4+ T cell activation and subsequent differentiation.*

Activation signals (1-3) of naïve CD4+ T cells are shown, followed by differentiated CD4+ effector/regulatory cells. Red text shows polarising cytokines (signal 3). Yellow text shows the hallmark transcription factors of each subtype. Blue text shows pro-inflammatory effector cytokines expressed by each effector subtype. Green text shows anti-inflammatory cytokines expressed by each regulatory/effector subtype.

Recently, research into non-coding RNAs (ncRNAs) as cellular regulators has provided compelling results detailing long ncRNAs (lncRNAs) as regulators of CD4+ T cell activation and differentiation (West & Lagos, 2019) (Plasek & Valadkhan, 2021) (Liu, et al., 2022).

## 1.2 Non-coding RNA

NcRNAs are RNA transcripts which are not translated into proteins and are classified into small (sncRNAs, 10-200nts long) and large (lncRNAs, over 200nt long) subcategories. Only 2% of the human genome is made up of protein coding genes, whilst up to 80% of it is transcribed into various types and subtypes of ncRNAs (The ENCODE Project Consortium, 2012). This discovery, along with research showing ncRNAs as important molecular regulators that can act through transcriptional, post-transcriptional and translational control (Yao, et al., 2019), has allowed a greater appreciation of ncRNAs as important cellular regulators and drivers of

disease, akin to that of protein. Most ncRNAs elicit their regulatory function through bound effector proteins (RNPs) and/or use their complementary antisense elements to interact with and regulate target nucleic acids (Matera, et al., 2007). Compared to protein coding genes, ncRNA genes are commonly poorly conserved, expressed at much lower levels and have greater tissue specific expression (Derrien, et al., 2012). As of July 2023, 27,488 ncRNA genes have been discovered within the human genome (lncRNA: 19922, sncRNA: 7566, GENCODE v44, July 2023). However, characterisation and detailed functional classification of these genes and their non-coding transcripts is severely lacking and requires further research (Zhao, et al., 2016).

### 1.2.1 Small non-coding RNAs

Examples of sncRNAs include small nuclear (snRNAs), micro RNAs (miRNAs) and transfer RNAs (tRNAs). tRNAs are integral for the translation of mRNA codons into appropriate amino acids for correct protein synthesis (Moore & Steitz, 2011). snRNAs associate with a protein complex, forming small nuclear ribonucleoproteins (snRNPs) within the nucleus. Within these snRNPs, snRNAs confer pre-mRNA binding through complementary base interactions whilst directing bound splicing machinery for regulation of nuclear pre-mRNA processing (Morais, et al., 2021). miRNAs were discovered as short (~22 nucleotides in length) transcripts which post-transcriptionally regulate gene expression through binding to complementary elements of mRNA in the cytoplasm (Wightman, et al., 1993). Traditionally, miRNA binding to complementary mRNA occurs in the 3'UTR and recruits the miRNA induced silencing complex (miRISC) to degrade the mRNA and silence gene expression. More recent research has shown that miRNA can bind anywhere along the transcript (ie. 5' UTR and coding sequences) and can also function as a translational activator; thus showing miRNAs are highly diverse in their function as post-transcriptional regulators (O'Brien, et al., 2018). For example, one of the most well-characterized miRNAs, miR-21, has a diverse range of mRNA targets which has been studied across various biological contexts, including cancer, immunology, and development (Krichevsky & Gabriely, 2009).

### 1.2.2 Long non-coding RNAs

Long non-coding RNAs are classified based on their transcription loci relative to protein coding genes within the genome. Intronic ncRNAs are transcribed between the exons of protein coding genes. Sense and antisense ncRNAs are transcribed from regions overlapping a part or

the whole length of a protein coding gene on the sense or antisense DNA strand respectively. Intergenic lncRNAs (lincRNAs) transcribed from genomic regions in between protein coding genes (Ma, et al., 2013). Intronic, sense and antisense lncRNAs, are less common than lincRNAs due to the large percentage of the genome being intergenic sequences and therefore lincRNAs represent the best characterised classification of lncRNAs. LncRNAs have a range of functions including regulation of transcription, translation and splicing. They can act in trans, affecting genes without direct interaction with their loci, or in cis, regulating genes through direct interactions with their locus, for example genes that are adjacent or overlapping with the lncRNA locus. Antisense lncRNAs, have gained particular interest for their potential to act as antisense cis regulators for their protein coding partner transcripts due to the fact that ~32% of the human lncRNAs are transcribed antisense to coding genes (Ma, Bajic and Zhang 2013) (Derrien, et al. 2012).

*HOTAIR*, a prominent trans-acting antisense lncRNA transcribed from the *Hoxc* locus, governs transcriptional regulation of the *Hoxd* gene cluster through interactions with chromatin modifying complexes to induce epigenetic changes (Heubach, et al., 2015). Dysregulation of *HOTAIR* has consequences impacting immune system development and cancer progression (Botti, et al., 2019). *Xist,* one of the best characterised cis-acting lincRNAs, is intrinsic for survival through dosage compensation of the X chromosome in females via facilitation of X inactivation (Marahens, et al., 1997). Dysregulation of *Xist* can lead to the abnormal expression of genes on the previously silenced X chromosome, resulting in the occurrences of X-linked diseases (Agrelo & Wutz, 2010). Several other lncRNAs are involved in regulating or driving disease phenotypes (Zhang, et al., 2017) (Hall & Lekka, 2018), therefore their study is vital for a deeper understanding of molecular disease mechanisms and the subsequent development of novel therapies. However, unlike protein, only a handful of lincRNA transcripts are well characterised and there are currently no ncRNA targeting therapies despite research existing to stress the importance of lincRNAs.

### 1.2.3 LncRNAs are regulators of CD4+ T cells

LncRNAs are found within a wide range of molecular and cellular contexts, recently rising as important regulators of the immune response. Concerning T cell function, the lncRNA transcriptome displays major changes within CD4+ differentiation trajectories, as some lncRNAs are characteristic of a naïve state whilst others are intrinsic for driving an activated

and differentiated state (Ranzani, et al., 2015) (Plasek & Valadkhan, 2021). Two antisense lncRNAs, *Infg-AS1* and *Gata3-AS1* are specifically expressed in Th1 and Th2 linages respectively (Liu, et al., 2022). Infg-AS1 is epigenetically activated by T-BET mediated and its expression positively cis-regulates Th1-lineage-specific expression of the hallmark cytokine *Ifng* (Collier, et al., 2014). *Gata3-AS1* was shown the negatively cis-regulate *Gata3* expression by mediating the placing of repressive epigenetic markers immediately adjacent to its own locus. Therefore, repression of *Gata3-AS1* was shown to be vital for initiation of Th2 activation and linage specific cytokine expression (Gibbons, et al., 2018). Further research of emerging lncRNAs is vital for a better understanding the regulation of CD4+ T cell activation and differentiation for subsequent identification of therapeutic candidates for CD4+ T cell mediated pathologies (West & Lagos, 2019).

### 1.3 *Malat1* is a critical regulator within cellular stress and disease states

*Malat1* is a lincRNA with a length of ~6.7kb in mice and ~7kb in humans (Wilusz, et al., 2012). *Malat1* is specifically enriched in the nuclear speckles along with several subclasses of splicing proteins (Hutchinson, et al., 2007). Since its discovery, *Malat1* has distinguished itself from the thousands of other lncRNAs simultaneously being uncovered. This is due to its high conservation between mammalian species (stretch of >7000nts showing 69.3% homology between mice and human *Malat1*), ubiquitous expression and high abundance, which are all uncommon features of lincRNAs (Hutchinson, et al., 2007) (Ji, et al., 2003). Additionally, *Malat1* possess a rare structural motif known as a triple helix at its 3' end, which in most RNA transcripts, would usually be occupied by a poly-(A) tail (Wilusz, et al., 2012). The triple helix is responsible for the high nuclear abundance of *Malat1* as the stability of the structure protects from degradation by exonucleases (Brown, et al., 2012). The structure and indispensability of the triple helix for the high abundance of *Malat1* make this a targetable lncRNA for knockdown experiments with increased druggability (Abulwerdi, et al., 2019).

Initially scientists hypothesised that *Malat1* was a regulator of a highly conserved developmental or homeostatic process, due to its distinguishable features. However, three independent studies generated *Malat1* knockout mouse models and concluded that mice can develop and reproduce successfully with a complete lack of *Malat1* (Eißmann, et al., 2012) (Zhang, et al., 2012) (Nakagawa, et al., 2012). Further histological and molecular analysis

yielded no obvious phenotypic defects of *Malat1*$^{-/-}$ mice. These studies therefore concluded that *Malat1* had no detectable role in physiological development in mice.

### 1.3.1 *Malat1* is a driver of metastasis in a range of cancer types

Despite *Malat1* deletion showing no functional role within homeostasis and development, there is significant evidence that it elicits control in stress and disease conditions (Zhang, et al., 2017). *Malat1* was first discovered as a predictive biomarker of poor prognosis and increased metastatic occurrence in patients of non-small cell lung cancer (Ji, et al., 2003). *Malat1* dysregulation is common in cancers. For example, significant association of *Malat1* overexpression and increased metastasis occurrence is displayed in thousands of cancer patients of varying cancer types (Li, et al., 2018). Studies have demonstrated that *Malat1* drives metastasis *in vivo* in several cancer models, including the genetic breast cancer model MMTV-PyMT and both EBC-1 and A549 xenograft lung cancer models. In these studies targeted downregulation of *Malat1* within primary tumours with antisense oligonucleotides decreases metastatic burden (Arun, et al., 2015) (Gutschner, et al., 2013). Contrary to these studies, evidence also exists suggesting that *Malat1* instead supresses metastasis (Kim, et al., 2018). Overall all of these studies support the importance *Malat1* in disease progression. However, a complete map of the mechanistic action by which *Malat1* regulates metastasis has not yet fully been discerned. This displays that research is still required to uncover the exact role and action of *Malat1* within metastasis prior to assigning greater confidence in the therapeutic potential of the transcript.

### 1.3.2 *Malat1* is a regulator of Th cell phenotype and host immunity towards infection.

More recently, increasing evidence is being published supporting a vital regulatory role for *Malat1* within the immune system. For example, altered *Malat1* expression compared to healthy patients has been detected in sufferers of autoimmune and inflammatory conditions including Diabetic Retinopathy (DR) (Saumik, et al., 2018) Multiple Sclerosis (MS) (Masoumi, et al., 2019) and Systemic Lupus Erthymatosus (SLE) (Yang, et al., 2017). Patients of SLE, an autoimmune disorder characterised by hyperactive T cells, were shown to have *Malat1* overexpression in several immune populations, including that of T cells (Yang, et al., 2017) (Gao, et al., 2020). An association of *Malat1* dysregulation in T cell mediated pathology suggested a possible regulatory role in CD4+ T cell activation and effector Th cell function.

Whilst one study using mouse infection models of lymphocytic choriomeningitus virus (LCMV) proposed that *Malat1* is dispensable for CD4+ T cell development and immunity towards LCMV (Yao, et al., 2018), several other studies have specifically implied a critical regulatory role for *Malat1* in CD4+ T cell activation and differentiation. A recent study from our lab identified *Malat1* as a regulator of both initial CD4+ T cell activation and Th cell hallmark cytokine expression as a loss of *Malat1* leads to a more enhanced Th cell response to Th1 inducing infections *in vivo* (Hewitson, et al., 2020). Within this study, initial results from *in vitro* CD4+ differentiation assays using *Malat1* targeting LNA gapmers (antisense oligonucleotides) for *Malat1* knockdown showed that *Malat1* upregulates IL10, the anti-inflammatory cytokine, expression in Th1/2 cells through the transcriptional regulator MAF. Due to the critical role of IL10 in downregulating Th cell action to prevent immunopathology (Saraiva & O'Garra, 2010) the effect of *Malat1* was investigated within Leishmania donovani and Plasmodium chabaudi chabaudi AS (PcAS) *in vivo* chronic infection models. Results showed that *Malat1*$^{-/-}$ mice had increased parasite clearance and severe immunopathology respective to each infection compared to Wildtype (*WT)* mice. Furthermore, Th1 cells derived from infected *Malat1*$^{-/-}$ mice had significantly lower IL10 expression. *Il10*$^{-/-}$ mice have similar response to infection in the form of a lack of immunosuppression (Rennick, et al., 2019), providing further evidence that the *in vivo* results above are indeed a consequence of loss of *Malat1* mediated *Il10* expression. Overall, this study displayed that *Malat1* acts as a suppressor of immunity to Th1-inducing parasitic infections, specifically maintaining the equilibrium between effective pathogen clearance and severe immunopathology (Hewitson, et al., 2020). These findings are supported another study which showed that a *Malat1* depletion increased the frequency of effector Th1/Th17 subtypes whilst decreasing Treg frequency within non-polarising *in vitro* activation assays. This same study associated *Malat1* downregulation with enhanced neuroinflammation in MS patients (Masoumi, et al., 2019). Results from these studies support that *Malat1* is a critical regulator of immunity through controlling CD4+ T cell activation, affecting Th phenotype through altered cytokine expression. This is further supported as *Malat1* downregulation is a hallmark of Th1 and Th2 activation (Hewitson, et al., 2020) and dysregulation of *Malat1* expression dynamics is detected in patients of DR and SLE (Saumik, et al., 2018) (Masoumi, et al., 2019).

Importantly, the mechanism for how *Malat1* regulates CD4+ T cell activation is not understood beyond the downregulation of IL10 via MAF and therefore requires further investigation within specific Th subtypes and disease states.

## 1.4 *Malat1* mechanisms of action

Within the diverse range of disease contexts that *Malat1* has been found to regulate, the molecular mechanisms of action of the non-coding transcript are complex and varied. Ongoing research aims to shed light on new regulatory mechanisms of *Malat1* as well as push better characterisation of the diverse range of known mechanisms. These include binding actively transcribed genes, polycomb repressive complex 2 (PCR2) interaction, action as a miRNA sponge, RBP interaction and regulating nuclear speckle organisation.

### 1.4.1 Transcriptional, post-transcriptional and translational regulatory mechanisms of *Malat1*.

Studies suggest that *Malat1* directly regulates transcription through identification of a plethora of trans-located chromatin binding sites where *Malat1* binds to active transcription sites (Engreitz, et al., 2014) (West, et al., 2014). *Malat1* can also act as a regulator of gene transcription via epigenome modelling through interaction and direction of PRC2 mediated tri-methylation of H3K27 (Chen, et al., 2020). The result of this epigenetic marker is the silencing of adjacent genes. Therefore, *Malat1* can either activate or inhibit gene expression by sequestering or promoting PRC2 presence at target genes. LncRNAs binding to and recruiting PRC2 to epigenetically silence genes is common; for example, being used in X activation by *Xist* (Loda & Heard, 2019). This mechanism has mainly been shown in the context of *Malat1* promotion of cancer progression, as *Malat1* knockout causes disassociation of PCR2 from tumour suppressor genes, preventing their silencing (Chen, et al., 2020) (Wang, et al., 2016). However, *Malat1* direction of PRC2 to regulate the epigenome of target genes has been shown in other contexts such as driving the oxidative stress response in skeletal muscle (El Said, et al., 2021) and driving HIV-1 transcription and replication in CD4+ T cells (Qu, et al., 2019), displaying the diversity of *Malat1* and PRC2 interaction.

LncRNAs, including *Malat1*, can function as post-transcriptional regulators by acting as miRNA sponges that bind and inhibit/sequester miRNAs, leading to a recovery of the effect elicited by the miRNA (Salmena, et al., 2011). This mechanism represents the antithesis of miRNA post-transcriptional regulation of mRNA transcripts, exhibiting complex layers of regulatory

crosstalk of different RNA transcripts. Research has mapped out a detailed network of cancer-related miRNA targets which *Malat1* is able to bind and sequester, causing a downstream effect on several oncogenes target expression (Su, et al., 2021). For example, in colon cancer *Malat1* inhibits expression of *miR-21*, leading to reduction in migration and invasion of colon cancer cells (Huang, et al., 2020). *Malat1* has also been shown to act as a miRNA sponge for *miR-150-5p*, which leads to increased apoptosis and extracellular matrix degradation in osteoarthritis (OR) a disorder linked to increased immune inflammation (Zhang, et al., 2019).

As well as regulation at the transcriptional and post-transcriptional level another notable mechanism of action of *Malat1* include binding to and regulating proteins factors which regulate translation, leading to increased translation of oncogenes including *Runx2* and *Tcf7l2* (Ji, et al., 2019) (Malakar, et al., 2019).

### 1.4.2 *Malat1* regulates organisation of nuclear speckles through riboregulation of SR proteins

*Malat1* is localised specifically within the nuclear speckles (Hutchinson, et al., 2007) (Spector & Lamond, 2011). Nuclear speckles are subnuclear compartmentalised organelles not bound by any membrane and have the function of processing and modifying pre-mRNA. As well as *Malat1*, these organelles are also highly enriched with splicing factors, including snRNPs and serine and arginine rich (SR) proteins (Cáceres, et al., 1997). Nuclear-speckle associated splicing factors, along with a multi-component ribonucleoprotein machinery, known as the spliceosome, localizes alongside pre-mRNA at nuclear speckle peripheries to induce and regulate pre-mRNA processing and splicing (Liao & Regev, 2021) (Hall, et al., 2006). Several proteomic studies show direct *Malat1* interaction with nuclear-speckle associated factors such as SR proteins, notably including that of SRSF1 (Scherer, et al., 2020) (Spiniello, et al., 2018) (Tripathi, et al., 2010).

An *in vitro* study within HeLa cells showed that *Malat1* depletion leads to a decreased association of a subset of pre-mRNA splicing factors to the nuclear speckles, including SRSF1, SRSF3 and U2snRNP (Tripathi, et al., 2010). This study also showed that *Malat1* directly bound to SRSF1 and that *Malat1* depletion caused a change in SRSF1 phosphorylation, a process shown to regulate SRSF1 alternative splicing function and localisation within the nucleus (Stamm, 2008) (Lai, et al. 2000). This suggests that lower *Malat1* binding to SRSF1 alters the functional capacity to act as a splicing factor due to an altered phosphorylation state and localisation away from the nuclear speckles. Results also showed that incorrect nuclear

speckle assembly caused by *Malat1* depletion lead to aberrant cell mitosis and alternative splicing (Tripathi, et al., 2010). This study assigned *Malat1* as an essential factor in the correct assembly and function of nuclear speckles, mainly concerning the localisation and activation state of nuclear speckle associated SR protein, specifically that of SRSF1.

RNP formation by RBPs binding to a specific RNA transcript is a common mechanism for regulating the isoform usage and abundance of the bound RNA transcript through alternative splicing and altered transcript stability (Quattrone & Dassi, 2019). As shown above with *Malat1* regulating SRSF1 phosphorylation state and localisation, this form of RNP-mediated regulation be inversely functional, with the RNA transcript regulating the function of its bound RBP partner (Tripathi, et al., 2010) (Hentze, et al., 2018). This method of RNA-mediated RBP regulation is known as 'riboregulation'. Another example of *Malat1* riboregulation of SRSF1 is shown as *Malat1* depletion in U2OS cell line leads to reduced SRSF1 localisation at a transgene locus (Benard, et al., 2010).

Of the range of *Malat1* regulatory mechanisms, riboregulation by *Malat1* will make up the scope of this study as we will specifically consider the riboregulation of nuclear speckle localised splicing factors bound by *Malat1* and investigate the effect of this on CD4+ T cell differentiation.

## 1.5 The SR protein family

The SR protein family of proteins represents a diverse and vital group of RNA-binding proteins found within the nuclear speckles (Cáceres, et al., 1997). These proteins share two domain types that characterise them as SR proteins. Firstly, the RNA recognition motif (RRM) which confer the ability to interact with pre-mRNA transcripts, making the SR proteins a group of RBPs. Secondly, all SR proteins contain an RS domain, enriched in serine and arginine amino acids, which is vital for conferring protein interactions with other RS domain containing splicing factors (Kohtz, et al., 1994). These two domains on all SR proteins act as critical components that facilitate precise RNA splicing metabolism by bridging interactions between pre-mRNA molecules and various splicing factors (Shepard & Hertel, 2009). Of the 12 characterised SR proteins (Howard & Sanford, 2016), SRSF1 stands the archetypal and most extensively studied example. SRSF1 is a multifaceted splicing factor with a diverse range of characterised interacting transcripts (Sanford, et al., 2009), including *Malat1* (Spiniello, et al.,

2018) and several other important immunological targets (Qi, et al., 2021) (Paz, et al., 2021). SRSF1 has been reported preferentially bind within purine-rich, exonic sequences (Wang, et al., 2005) (Sanford, et al., 2009). Since its discovery it has been cemented as a central regulator for dynamic control of gene expression and a key factor in cellular processes and development of disease.

### 1.5.1 SRSF1 mechanisms of action

SRSF1 binding to RNA transcripts via its RRM can regulate the transcript in several ways. Through its action as an alternative splicing regulator, SRSF1 can alter the relative ratios of different resultant isoforms of the same pre-mRNA. SRSF1 is also capable of regulating mRNA abundance through binding mRNA UTRs and regulating transcription through active recruitment to chromatin sites (Paz, et al., 2021) (Moulton, et al., 2013). Evidence also exists showing that SRSF1 is also capable of regulating its targets, such as those involved in cell cycle progression, outside of the nucleus at the translational level (Maslon, et al., 2014). These SRSF1 functions have been displayed in an immunological context and the direction of its action are extremely transcript dependent. This is shown by increased SRSF1 binding respective to each adjacent pre-mRNA site promoting exon exclusion for *CD6* Exon 5 (Gloria, et al., 2014), whilst inhibiting exon exclusion in *CD45* Exon 5 due to competitive binding of CD45 alongside the splicing factor hnRNPL (Motta-Mena, et al., 2010). These findings emphasise the multifaceted range of mechanisms that SRSF1 can exert to regulate a complex network of RNA transcript targets.

### 1.5.2 SRSF1 is a negative regulator of CD4+ T cell activation and differentiation

Analysis of SRSF1 within the immune system led to findings displaying SRSF1 as a regulator of T cell activation and phenotype. SRSF1 was found to be essential for thymocyte development and viral clearance in mice, with its deletion leading to a late-stage thymocyte maturation block (Qi, et al., 2021) and lower proportions of cytotoxic CD8+ T cells upon LCMV infection (Juarez, et al., 2022). Moreover, mice with a conditional knockout of *Srsf1* (*Srsf1*-cKO mice) in mature T cells, displayed more hyperactive T cells and SLE-like disease under non-infected conditions (Katsuyama, et al., 2019). CD4+ T cells of *Srsf1*-cKO mice were characterised by higher proportion of pro-inflammatory cytokine expression with a greater proliferative capacity. *Srsf1*-cKO mice also displayed severe signs of autoimmunity and lupus-like inflammation, characterised by higher levels of autoantibodies in the periphery and increased

T cell infiltration into internal organs. Another study using *Srsf1*-TregKO mice, which had Treg-specific SRSF1 knockout, showed that SRSF1 is essential in Treg function. *Srsf1*-TregKO mice succumbed to fatal systemic autoimmune disease and showed Tregs developing a much more pro-inflammatory phenotype, alike to effector Th cells (Katsuyama & Moulton, 2021). Overall, these studies suggest a role for SRSF1 as a negative regulation of the activation, differentiation and cytokine expression CD4+ T cells subtypes, similar to that described previously for *Malat1*. A controlled and directed reduction of SRSF1 abundance upon CD4+ T cell activation (Gloria, et al., 2014), is likely responsible for an appropriate effector Th response towards infection. An aberrant downregulation of its levels, therefore, tips the balance too far towards an effector phenotype leading to systemic autoimmunity and T cell hyperactivity. This hypothesis of SRSF1 regulation within CD4+ T cells is supported by human studies. Patients of SLE, an autoimmune disease characterised by hyperactive T cells, have lower expression of SRSF1 in T cells compared to healthy individuals (Kono, et al., 2018) (Moulton, et al., 2013). Whilst some targets of SRSF1 are well characterised within the immune system, a complete mechanism involving any transcriptional, post-transcriptional or translation targets of SRSF1 for regulating CD4+ T cell activation has not yet been established.

## 1.6 My Project

*Malat1* and SRSF1 have both emerged as critical regulators of CD4+ T cell activation and subsequent differentiation (Hewitson, et al., 2020) (Katsuyama, et al., 2019). Their significance is underscored by their direct associations across a multitude of biological contexts (Tripathi, et al., 2010) (Spiniello, et al., 2018). This includes that of recent work within the Lagos lab by Katie West which has shown through RNA antisense purification coupled with mass spectrometry (RAP-MS) (McHugh & Guttman, 2018) that *Malat1* and SRSF1 interact within primary CD4+ T cells (unpublished data). *Malat1* is able to affect the localisation and phosphorylation state of SRSF1 upon direct interaction, leading to altered transcriptional and post-transcriptional action by the SR protein (Tripathi, et al., 2010) (Benard, et al., 2010). Therefore, it is realistic to speculate that this regulatory axis is contributing to the molecular action exerted by each binding partner across the course of CD4+ T cell activation to endpoint Th/Treg differentiation.

*Malat1* loss has a more marked inhibitor effect on IL10 expression in Th2 cells compared to within Th1 cells (Hewitson, et al., 2020). IL10 expression, whilst important in all effector

subtypes, is a more characteristic and prominent feature of the Th2 subtype (Saraiva & O'Garra, 2010) Additionally, SRSF1 knockdown causes downregulation of IL4, a vital Th2 effector cytokine (Katsuyama, et al., 2019). Whilst SRSF1 and *Malat1* have been shown to be involved in regulation of several CD4+ T cell subtypes, this evidence emphasises the importance of these factors on regulating the correct outcome of Th2 phenotype. Therefore, this subtype will make up the main investigative route of this project.

### 1.7 Project Hypothesis and Aims

We hypothesised that *Malat1* regulation of Th2 cell differentiation is mediated at least in part through riboregulation of SRSF1 binding and resultant indirect transcriptional and/or post-transcriptional regulation of CD4+ T cell relevant RNA targets.

We aimed to investigate whether *Malat1* loss affected SRSF1 targets and function in Th2 cells using iCLIP. Following this, we aimed to test regulation of transcript usage and abundance of candidate transcripts by both *Malat1* and SRSF1. In addition to already available methods of *Malat1* knockdown/knockouts we aimed to establish SRSF1-deficient CD4+ T cell models through use of CRISPR-Cas9.

Individual aims were as follows:

1. Characterise the alternative RNA binding behaviour of SRSF1 upon *Malat1* loss in Th2 cells.
2. Test abundance and transcript usage of selected candidates in *Malat1*$^{-/-}$ Th2 cells.
3. Knockout SRSF1 using the CRISPR-Cas9 system, initially in the EL4 cell line and followed by primary naïve CD4+ T cells to validate the effect of SRSF1 on selected targets.

## 2. Materials and Methods

### 2.1 iCLIP

*In vitro* activated female Th2 cells (*wildtype* and *Malat1*$^{-/-}$) were prepared in the Lagos lab by Katie West and sent to the Ule Lab (Francis Crick Institute) at 6-days post activation for individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP) to be performed according to Hupertz, et al. Here, cells were irradiated with ultraviolet light (254nm) to permanently crosslink all bound RNA and protein within these Th2 cells. Following this, cells were lysed and RNA was partially fragmented to ease subsequent protein

immunoprecipitation. SRSF1, along with all crosslinked RNA, was immunoprecipitated through incubation with an SRSF1 targeting antibody conjugated magnetic beads. Remaining RNA from the precipitated sample was then dephosphorylated and ligated to an L3 linker, enabling subsequent amplification. RNA was also radioactively labelled at this timepoint. Immunoprecipitated samples were then ran on an SDS-page gel and transferred to a nitrocellulose membrane, where SRSF1 crosslinked RNA was specifically identified by an autoradiograph and subsequently removed from the membrane. Isolated SRSF1 crosslinked RNA underwent reverse transcription to produce cDNA, which was then circularised, amplified and sequenced to provide a genome-wide overview of RNA crosslinked to SRSF1 in *wildtype* and *Malat1*$^{-/-}$ Th2 cells. Computational processing of cDNA sequences allowed for single nucleotide resolution of SRSF1-RNA crosslink sites due to identification of truncation points where SRSF1 bound RNA transcripts and halted reverse transcription.

## 2.2 Cell Culture

EL4 cells were cultured in sterile conditions in Dulbecco's modified Eagle's Medium (DMEM) (Gibco, 11965092) supplemented with 1% Penicillin/Streptomycin (Gibco, 15140122), 10% Fetal Calf Serum (FCS) (Sigma-Aldrich, F2442) and 1% L-glutamine (Gibco, 25030081), at 37°C in 5% $CO_2$. All subsequent mention of DMEM refers to the aforementioned supplemented version. DMEM was replaced concurrently with a 1:10 dilution of EL4 cells every 72-96 hours or after use in an experiment, maintaining the culture at around 10ml. For long-term storage of EL4 cell populations, $1x10^6$ cells were resuspended in 10% DMSO (PanReac AppliChem, A3672), 90% FCS and placed in −196°C liquid nitrogen. To retrieve EL4 cells frozen in liquid-nitrogen, samples were gently returned to room temperature and resuspended in DMEM.

## 2.3 I*n vitro* CD4+ T cell activation

### 2.3.1 Sample collection

Spleens and lymph nodes were collected from *wildtype* C57BL/6 strain mice and *Malat1*$^{-/-}$ mice (Nakagawa, et al., 2012). All mice used were female and bred in-house under specific pathogen free conditions. Harvested spleens and lymph nodes were placed in DMEM and kept on ice. All subsequent steps involved processed samples being kept on ice where possible.

### 2.3.2 CD4+ T cell isolation

Harvested spleens and lymph-nodes together were separated into a single cell solution with up to 15ml of DMEM through Falcon™ Cell Strainers (Falcon, 352350). Cells were resuspended (all further resuspensions using primary samples occurred at: 1500rpm, 4°C for 5 minutes) in 2ml ACK buffer (Gibco, A10492) to lyse red blood cells. Following this, cell samples were further resuspended in 1ml of DMEM and a 1in20 dilution was carried out to determine total non-red blood cell harvest number using Trypan Blue to stain live cells and a haemocytometer.

Per $1x10^7$ cells, cell samples were resuspended in 40µl of MACS (Miltenyi Biotec, 130-091-221) buffer and 3µl biotin-conjugated antibody cocktail (Miltenyi Biotec,130-104-454) for negative selection of CD4+ T cells and incubated for 10 minutes at 4°C. A further 20µl of MACS buffer as well as 6µl of Anti-Biotin MicroBeads (Miltenyi Biotec,130-104-454) were added per $1x10^7$ cells. A further 15-minute 4°C incubation ensued, followed by a wash in MACS buffer and resuspension in 1ml of MACS buffer. These samples were then run through pre-rinsed magnetic LS columns (MACS, 130-042-401). Negatively selected CD4+ T cells outputted from this step were quantified using Trypan Blue and a haemocytometer.

### 2.3.3 Naïve CD4+ T cell activation with Th2 polarisation conditions

Prior to stimulation of purified CD4+ T cells, a flat-bottom 96-well plate was prepared with the appropriate number of wells (1 well per $5x10^5$ cells) through addition of 50µl PBS (Gibco, 10010023), containing 10 mg/ml anti-CD3 (Biolegend, 145C11). Following incubation of this plate at 37°C for 4 hours, the contents of wells was removed, and wells were washed gently in PBS. 200µl of purified CD4+ T cells (resuspended in DMEM at $5x10^5$ cells per 200µl) was the added to each anti-CD3 coated well. Additionally, 4µg/ml of anti-CD28 (Biolegend, 102105) was added as the co-stimulatory signal alongside 30 ng/ml mouse rIL-4 (PeproTech, 214-14-1MG) and 5 µg/ml anti–IFNγ (Biolegend, 505807) to specially induce Th2 polarisation. Stimulated CD4+ T cells were incubated at 37°C.

Four days post-activation CD4+ T cells were taken from the plates for counting (following 1:4 dilution with DMEM) and two DMEM washes. Cells were then re-plated in a 96 well plate at a volume of 200µl along with 10ug/ml of rIL-2 (PeproTech, 200-02-1MG) and incubated at 37°C. Cells were removed for analysis by flow cytometry, RNA quantification or western blotting at day 4 (prior to rIL-2 treatment), day 5 or day 6 post activation signal introduction.

## 2.4 CRISPR-Cas9 genome editing

### 2.4.1 CRISPR-Cas9 RNP production

CRSIPR-Cas9 mediated genome editing was carried out using The Alt-R™ CRISPR-Cas9 System (IDT), specifically using reagents including HPRT targeting crRNA (IDT, 1072541), Negative Control crRNA (IDT, 1072544) and tracrRNA (IDT, 1072532) from the mouse CRISPR-Cas9 Control Kit (IDT, 1072555). Also used were Alt-R™ *S. pyogenes* Cas9 Nuclease V3 (IDT, 1081058) and pre-designed *Srsf1* targeting crRNAs, SRSF1-AA (IDT, Mm.Cas9.SRSF1.1.AA) and SRSF1-AC (IDT, Mm.Cas9.SRSF1.1.AC). BLAST was used to confirm the intended *Srsf1* mRNA target sequence in exon 1 of these crRNAs. Targeted sequences within the *Srsf1* gene were selected according to their loci being immediately proceeded by the correct protospacer adjacent motif (PAM) – NGG, which is essential for the correct binding and double stranded DNA cleavage by *S. pyogenes* Cas9 (Guo, et al., 2019). The sequences of crRNA are found in Table 4. The sequences of HPRT and negative control crRNA were not disclosed by manufacturers. CRISPR-Cas9 reagents were kept on ice wherever possible.

To make tracrRNA-crRNA duplexes equal volumes of 200µM stock solutions of crRNA and tracrRNA were used. To anneal this solution into a suitable guide RNA (gRNA) it was incubated at 95°C for 5 minutes. Cas9 nuclease was then added to gRNA and incubated at room temperature for 20 minutes to form CRISPR-Cas9 RNPs. In EL4 cell experiments, an RNP concentration of 1µM was used, with 1µl of 200µM crRNA (0.5µl each of SRSF1-AA and SRSF1-AC crRNA in double guide treated conditions), 1µl of 200µM tracrRNA and 1.6µl of 62µM Cas9 per 100µl electroporation reaction. For Th2 experiments, cells a final RNP concentration of 2µM was used, with gRNA in a ~3-fold excess compared to Cas9 nuclease. 2µl of 200µM crRNA (1µl each of SRSF1-AA and SRSF1-AC crRNA in double guide treated conditions) and 2µl of

*Table 1.* crRNA sequences used in CRISPR-Cas9 RNPs for electroporation mediated transfection

| crRNA | Target | Target Sequence | PAM sequence |
|---|---|---|---|
| SRSF1-AA (Mm.Cas9.SRSF1.1.AA) | *Srsf1* | TATCCGAACCAAGGACATCG | AGG |
| SRSF1-AC (Mm.Cas9.SRSF1.1.AC) | *Srsf1* | AACGACTGCCGCATCTACGT | AGG |

200µM tracrRNA was used alongside 1µl of 62µM Cas9 nuclease per 30µl electroporation reaction.

### 2.4.2 EL4 cell line electroporation mediated CRISPR-Cas9 RNP transfection

This method was developed following optimisation and relates directly to the viability seen in Figure 17c and knockout efficiency seen in Figure 18b. Per electroporation reaction, $1x10^6$ EL4 cells were taken from DMEM and washed in PBS three times (1500rpm, 10mins, 4°C), finally being resuspended in 100µl of PBS along with annealed RNP complexes. Following gentle mixing and a 10-minute incubation at 4°C, EL4 cells and RNP complexes were transferred into 0.2cm gap Gene Pulser Electroporation Cuvettes (BioRad, 1652082). These were electroporated at 220V for two 2ms pulses in the Gene Pulser Xcell system (BioRad, 1652660). Immediately after electroporation, ice cold DMEM was added the contents of the microcuvette and left to gently reach room temperature. Electroporated cells in DMEM were then placed into pre-warmed DMEM media in 24 well plates and incubated at 37°C.

Following electroporation of EL4 cells with SRSF1-AA and/or SRSF1-AC crRNA associated RNPs, cloning by limiting dilution was carried out. Cells were made up to a concentration of $5x10^3$ in 200µl of DMEM. This was added to well A1 of a flat bottomed 96 well plate and diluted at a factor of 2 down the 1st column of the plate. Using a P200 multichannel pipette diluted cell samples were diluted by further factor of 2 across each subsequent column. This created a diagonal dilution gradient across the 96 well plate, in the attempt of diluting certain wells to a single cell per 200µl of DMEM, leading to a clonal EL4 population through expansion. Plates were incubated at 37°C for several weeks, with regular monitoring of wells with initially small and expanding populations. Expanded populations were removed and placed in warmed DMEM media in 48 well plate. Populations that expanded further from this stage were then placed in warmed DMEM in a 24 well plate for monitoring of SRSF1 expression by western blotting. Following screening of initial expanded populations, a second round of cloning by limiting dilution was carried out with the same methodology as above.

### 2.4.3 Naïve CD4+ T cell electroporation mediated CRISPR-Cas9 RNP transfection

This method of transfection was adapted from previous published experiments (Seki & Rutz, 2018). Its aim was to knockout a target prior to *in vitro* CD4+ T cell activation; therefore CRISPR-Cas9 RNP transfection was carried out following section 2.3.2 (CD4+ T cell isolation). Isolated naïve CD4+ T cells were resuspended in P3 Primary Cell Nucleofector™ Solution

(Lonza, V4XP-3032) at a concentration of $3x10^6$ cells per 25µl. 25µl of resuspended naïve CD4+ T cells were then added to each well of a Nucleocuvette® strip (Lonza, V4XP-3032) along with 5µl of RNP complex. Following 10 minutes of incubation at 4°C the Nucleocuvette® strip was electroporated with setting 'DS137' on the 4D-Nucleofector® X Unit (Lonza, AAF-1003X). Immediately after electroporation cold DMEM was added to nucleofector cuvette strips and gently returned to room temperature. Electroporated naïve CD4+ T cell content of each cuvette was made up to 300µl using warmed DMEM. *In vitro* CD4+ T cell activation was continued following this electroporation step into section 2.3.3 (Naïve CD4+ T cell activation with Th2 polarisation conditions) with the only modification of $1x10^6$ input cells (100µl of each 300µl cuvette output) per anti-CD3 coated well.

## 2.5 Gapmer mediated SRSF1 knockdown in EL4 cells.

An *Srsf1* targeting Antisense LNA gapmer (Qiagen, LG00784568-DDA) was used in EL4 cells to assess their capability to reduce *Srsf1* mRNA abundance compared to a Negative control Antisense LNA gapmer (Qiagen, LG00000002). All gapmers were used at a final concentration of 0.1µM. For gymnotic uptake cells were seeded at $2.5x10^5$ per 475µl of DMEM in wells of a 24 well plate. Following this 24µl of Opti-MEM reduced serum media (Gibco, 31985062) containing 1µl of 50µM stock gapmer was added to cells. For lipofection, cells were seeded at $2.5x10^5$ per 450µl of DMEM in wells of a 24 well plate. 1µl of Lipofectamine™ 2000 Transfection Reagent (Invitrogen, 11668019) was added to 24µl of Opti-MEM and vortexed thoroughly. Lipofectamine™ 2000 containing Opti-MEM was then added to a further 24µl of Opti-MEM with 1µl of 50µM stock gapmer, gently mixed and incubated at room temperature for 10 minutes. Following this the Lipofectamine™ 2000 and gapmer solution was added to the EL4 cell plate. For electroporation mediated gapmer transfection, EL4 cells were washed 3 times in PBS and seeded at $2.5x10^5$ EL4 cells in 75µl of PBS. This was transferred to a 0.2cm gap Electroporation Cuvettes along with 24µl of Opti-MEM and 1µl of gapmer. Electroporation was then carried out at the defined condition (200V, 1 pulse or 200V, 2 pulses) in the Gene Pulser Xcell system. Following this immediate addition of 100µl of cold DMEM into electroporated EL4 cells and gentle return to room temperature occurred. Cuvette contents were added to 400µl of warmed DMEM a in 24 well plate. All gapmer treated EL4 cells were transferred into an incubator at 37°C for 72 hours, after which their SRSF1 expression was assessed by qPCR and western blotting.

## 2.6 RNA quantification

### 2.6.1 RNA extraction

Extracted cells (EL4/Th2) were transferred to a 2ml Eppendorf for two PBS washes (1500rpm, 10mins, 4°C). Supernatant was removed and 700µL of QIAzol Lysis Reagent (Qiagen, 79306) and 140µl of Chloroform was added to cell pellets and mixed vigorously, incubated for 5 minutes, then centrifuged (12,000G, 15mins, 4°C). The upper clear aqueous phase was transferred into a new eppendorf and mixed with 450µl of 100% ethanol. Extracted RNA was then washed by centrifugation following transfer into RNeasy® Mini columns (Qiagen, 74124). Washes were as follows: 650µl of RWT (Qiagen, 1067933) (10,000G, 15sec, 4°C), 450µl of RPE (Qiagen, 1018013) (10,000G, 15sec, 4°C), 450µl of RPE (10,000G, 2min, 4°C), Dry wash (10,000G, 2min, 4°C). The top RNA containing section of the RNeasy® Mini column was inserted into a separate collection eppendorf and 30µl of Nuclease Free water (Qiagen, 129115) was added directly to the centre of the filter. Columns were spun (10,00G, 1min, 4°C) to collect 30µl of purified RNA solution which was stored at -80°C.

### 2.6.2 Reverse transcription/ cDNA synthesis

cDNA synthesis was carried out using two master mixes (MM). MM1 contained the following per 5ul RNA sample: 1µl 50ng/µl Random Hexamer Primers (Promega, C1181), 1µl 10nM dNTPs (Promega, U1511) and 5.5µl of nuclease free water. MM2 contained the following per 5ul RNA sample: 4µl 5X First-Strand Buffer, 2µl 100mM DTT, 0.5µl 200 U⁄µl SuperScript II Reverse Transcriptase (Invitrogen, 18064022) and 1µl of 100 mM RNaseOUT™ (Invitrogen, 10777019). Concentration of RNA was not controlled during reverse transcription, however later Ct results gained from qPCR were normalised to the housekeeping gene, U6. For reverse transcription 5µl of each RNA sample was added to 7.5µl of MM1 in PCR tubes and ran through a PCR thermal cycler machine (10 min at 25°C). 7.5µl of MM2 was added to each sample and the remainder of the PCR cycle carried out (50 minutes at 50°C, 5 minutes at 85°C) giving 20µl of cDNA from each sample which was stored at -20°C.

### 2.6.2 qPCR Primer design and validation

Primers were designed for this project using Primer Quest. Primer Blast was used to ensure isoform specificity of *Il2ra* and *Runx3* primers. All primer sequences used are detailed in Table 4. 6-day post-activation *wildtype* Th2 cell cDNA was diluted 1:4 to produce a serial dilution of 6 samples (1:1–1:256). These samples were used in qPCR to calculate the efficiency of designed primers, through the gradient of Ct values across increasing dilutions of cDNA with each primer pair. Primers with an efficiency of 90-110% and a single peak in melt curve (negative derivative reporter against temperature) was deemed acceptable for use in further experiments and inclusion in this report. U6 was used as a housekeeping gene and all primers were used at a working concentration of 10µM.

*Table 2.* Primers used for RNA quantification of targets in RTpPCR

| Target | Forward Primer Sequence (Sense) | Reverse Primer Sequence (Anti-Sense) |
|---|---|---|
| *Srsf1* | ACGCGGTGTATGGTCGCGAC | TGTTCCACGGCCGCTTCGAG |
| *Il2ra* | AAGCCAAGATGACAGACTGAG | CCCGACAACTGGGTACTATG |
| *Long Il2ra* | GCGTTGCTTAGGAAACTCCTGG | GCATAGACTGTGTTGGCTTCTGC |
| *Truncated Il2ra* | CCAGCCATTCATGCATTAGG | CTCATTGGCAGATGCAAGTT |
| *Runx3* | AGTGGGCGAGGGAAGAGTTTC | GCCTTGGTCTGGTCTTCTATCT |
| *Distal Runx3* | CAAAACAGCAGCC AACCAAGT | AGATGCTGTTGG AAGCCATGT |
| *Proximal Runx3* | CGTATTCCCGTAGACCCGAG | AGGGGAAGGCCGTGGAG |
| *U6* | CGCTTCGGCAGCACATATAC | TTCACGAATTTGCGTGTCAT |

### 2.6.4 Quantitative Real Time polymerase chain reaction

1µl of cDNA from each sample, along with 10µl of SYBR Green (Sigma-Aldrich, A46109), 0.6µl of each 1µM primer (forward and reverse) and 7.8µl of nuclease free water, was added to a well on a 96 well PCR plate. The StepOnePlus Real-Time PCR System (Applied Biosystems) was used to output the Ct values of individual Quantitative RT-PCR reactions following 40 PCR cycles (60-95°C). Ct values were processed into relative transcript levels using the ΔΔCt method.

## 2.7 Western Blotting

### 2.7.1 Sample collection and protein extraction

Cells (EL4/Th2) were washed with 3 times with PBS (1500rpm, 10mins, 4°C) and lysed with 30µl radioimmunoprecipitation assay (RIPA) buffer (Table 4) containing 1:100 diluted

protease and phosphatase inhibitors (Sigma-Aldrich, P8340, P5726, and P0044). Following this, centrifugation occurred (12,000G, 15mins, 4°C) and the protein containing supernatant was taken to be stored at -20°C.

### 2.7.2 Protein sample concentration quantification – BCA assay

Quantification of protein samples concentration was determined using a Pierce™ Bicinchoninic Acid (BCA) assay kit (Thermo Fischer Scientific, A55864). Unknown protein samples were tested against samples of known protein concentration (2000, 1000, 500, 250 and 0 ug/µl) generated from Albumin standards. 2µl of unknown Th2/EL4 protein sample was added to 10µl of PBS for a 1:6 dilution. 5µl of diluted protein sample/known protein sample was added to 93.1µl of BCA reagent A and 1.9µl of BCA reagent B in one well of a flat bottomed 96 well plate. These were incubated at 37°C for 30 minutes. The absorbance of samples was analysed with Infinite® MPlex plate reader at a wavelength of 562. Outputted values were normalised to the standards used to calculate the concentration (µg/µl) of undiluted protein samples.

### 2.7.3 Gel electrophoresis and membrane transfer

Acrylamide SDS-PAGE gels were made in house; stacking gel (5%) and resolving gel (15%) recipes can be found in Table 4. The 10µg of protein was made up to 20µl using distilled water (ddH2O). 5µl of 4x loading buffer (Table 4) was added to the protein sample, vortexed thoroughly and denatured at 95°C for 10 minutes. 5µl of PageRuler™ protein ladder (Thermo Fischer Scientific, 11852124) and 16µl of denatured protein sample was added into stacking gel wells and ran into resolving gel at 120V in SDS-Page running buffer (National Diagnostics, EC-868) for 90 minutes.

Prior to transfers, PVDF Membrane (65x80mm) (Sigma-Aldrich, IPVH00010) were activated for 1 minute in methanol and subsequently washed for 2 minutes in dd$H_2O$ and 5 minutes in Trans-Blot Turbo Transfer Buffer (BioRad, 10026938). Protein was transferred from Acrylamide SDS-PAGE gel onto a PVDF membrane with transfer buffer using a BioRad Trans-Blot Turbo Transfer System ran at 25V for 10 minutes. Following transfer, membranes were blocked in 2% Bovine Serum Albumin (BSA) (Sigma-Aldrich, A7030) in 1xTBST (Table 4) for 30 minutes enduring constant movement on a falcon tube roller.

### 2.7.3 Antibody incubation and chemiluminescent quantification

Protein membranes were incubated with primary antibodies on constantly moving falcon tube rollers either overnight at 4°C or for 2 hours at room temperature (specified in Table 4). Membranes were then washed three times in 1xTBST for 10 minutes under constant movement and subsequently incubated for 105 minutes in Horseradish peroxidase (HRP) conjugated anti-rabbit/anti-mouse secondary antibodies (Dako, P0448/P0447). Following another three TBST washes, membranes were covered with Amersham ECL Western Blotting Detection Reagent (Cytiva, 1059250, 1059243) for chemiluminescent protein band visualisation using a ChemiDOC imager. Images were saved in TIF format and analysed using ImageJ where blotted targets were compared against Histone H3 or Beta-actin (Table 4) loading controls.

*Table 3.* Primary Antibodies used for Western Blotting

| Target | Antibody | Suspension | Isotype | Incubation |
|---|---|---|---|---|
| SRSF1 | ProteinTech, 12929-2-AP | 1:1000, 5ml 1x TBST, 5% BSA | Rabbit | Overnight |
| Beta-Actin | Abcam, [AC-15] ab6276 | 1:5000, 5ml 1x TBST, 5% BSA | Mouse | 2-hours |
| Histone H3 | XP® (D1H2) #4499 | 1:2000, 5ml 1x TBST, 5% BSA | Rabbit | Overnight |

## 2.8 Flow Cytometry

### 2.8.1 Cellular staining

Th2 samples taken for analysis by flow cytometry at either 4-days (prior to IL-2 treatment), 5-days or 6-days post activation. Samples were initially stimulated for cytokine release within their culture plates by adding 10µg/ml Brefeldin (Sigma-Aldrich, B7651), 0.5µg/ml PMA (Sigma-Aldrich, I0364) and 1µg/ml ionomycin (Sigma-Aldrich, P8139) and incubating at 37°C for 4 hours. Following this, cell samples were harvested from the culture plates and placed on ice, with a small number of cells being set aside for control staining (isotype, live/dead, unstained and single stain controls). All samples were washed in PBS and then resuspended in 1ml of PBS. For all following fluorescent antibody staining and incubation samples were kept on ice and in the dark.

Initially 0.2µl of Zombie Aqua™ antibody (Biolegend, 423101) was added to all of the Th2 cell samples as well as the isotype and live dead control samples and incubated for 10 minutes.

After this, all samples were washed in 1ml of fluorescence associated cell sorting (FACS) (Table 4) buffer and resuspended by vortex in 3µl of 2mg/ml rat IgG (Sigma-Aldrich, SAB4600100), for a 5-minute incubation.

For surface staining, 0.4µl anti-CD4-PerCP (Biolegend, 100540) and 0.25µl anti-TCRb-PECy7 (Biolegend, 109207) was added to Th2 cell samples, the isotype control and the CD4 and TCRb single stain controls. All of these samples were vortexed and then incubated for 20 minutes at 4°C in the dark.

Following a wash and resuspension of all Th2 cell samples and controls in 1ml of FACS buffer, fixing of cells occurred through addition of 150µl of Cytofix for a 20-minute incubation. Following this all samples were washed twice in 500µl of saponin supplemented PERM buffer (BD Bioscience, 554723), for maintenance of membrane permeabilization. Samples from this point could maintained in a fixed and permeabilised state when left at 4°C in 500µl of PERM buffer for up to 48 hours.

For intracellular staining, 0.4µl IL10-PE (Biolegend, 505008) and 0.4µl IL4-APC (Biolegend, 504105) were added to Th2 cells samples as well as IL4 and IL10 single stain samples and incubated for 20 minutes. Samples were then washed twice more in PERM buffer and once more in FACS buffer before being resuspended in 200µl of FACS buffer.

### 2.8.2 Fluorescence quantification

Fluorescently stained samples were processed in the CytoFLEX V0-B5-R0 Flow Cytometer and results were analysed using CytExpert for CytoFLEX Acquisition and Analysis Software. Initially single stain controls were processed for auto-compensation of emission frequencies between different fluorophores within Th2 cell samples. Following this, Th2 cell samples were ran and recorded at a minimum of 10,000 events. Sequential gating was used to sequentially identify live populations of cells, singlets and CD4+TCRb+ cells. Following this, gating was carried out

within CD4+TCRb+ cell populations to identify percentage of IL4+ and IL10+ cells within this subpopulation.

*Table 4.* Recipes used for reagents/buffers made in house.

| Reagent/Buffer | Recipe |
|---|---|
| 1x TBST | 10x TBS (pH 8.0):<br>1l $ddH_2O$, 12.2g Tris HCl, 87.65g NaCl.<br>1x TBST:<br>900ul $ddH_2O$, 100ul 10x TBS, 500ul Tween 20 |
| RIPA buffer | 150mM 0.876g NaCl, 10mM 1mL 1M Tris-HCl (pH 7.2), 0.1% 1mL SDS, 0.1% 100µL Triton X-100, 1% 1g Sodium Deoxycholate, 5mM 1mL 0.5M EDTA |
| Stacking Gel (5%) | 3.4ml $ddH_2O$, 830ul 30% Acrylamide Mix, 630ul 1.0M Tris (pH 6.8), 50ul 10% SDS, 50ul 10% Ammonium Persulfate, 5ul TEMED |
| Resolving Gel (15%) | 4.6ml $ddH_2O$, 10ml 30% Acrylamide Mix, 5ml 1.5M Tris (pH 8.8), 200ul 10% SDS, 200ul 10% Ammonium Persulfate, 8ul TEMED |
| Western Blot 4x Loading Buffer | 250nM Tris HCL (pH 6.8), 8% SDS, 10% Glycerol, 5% β-Mercaptoethanol, 0.05% Bromophenol Blue |
| FACS Buffer | 1l PBS, 0.5% BSA, 0.05% Azide |

## 2.9 Statistics

All statistics were carried out using GraphPad Prism 10, any specific tests and their significance outputs are detailed in relevant figure legends within section 3 (Results).

# 3. Results

## 3.1 Analysis of alternative binding behaviour of SRSF1 in *Malat1*$^{-/-}$ Th2 cells

The RNA binding behaviour of SRSF1 is intrinsically linked to its function. Bound loci within target transcripts determine the functional output of SRSF1 on these transcripts in terms of alternative splicing or mRNA stability (Gloria, et al., 2014) (Paz, et al., 2021) (Qi, et al., 2021). To explore riboregulation by *Malat1* on the RNA binding behaviour of SRSF1 during CD4+ T

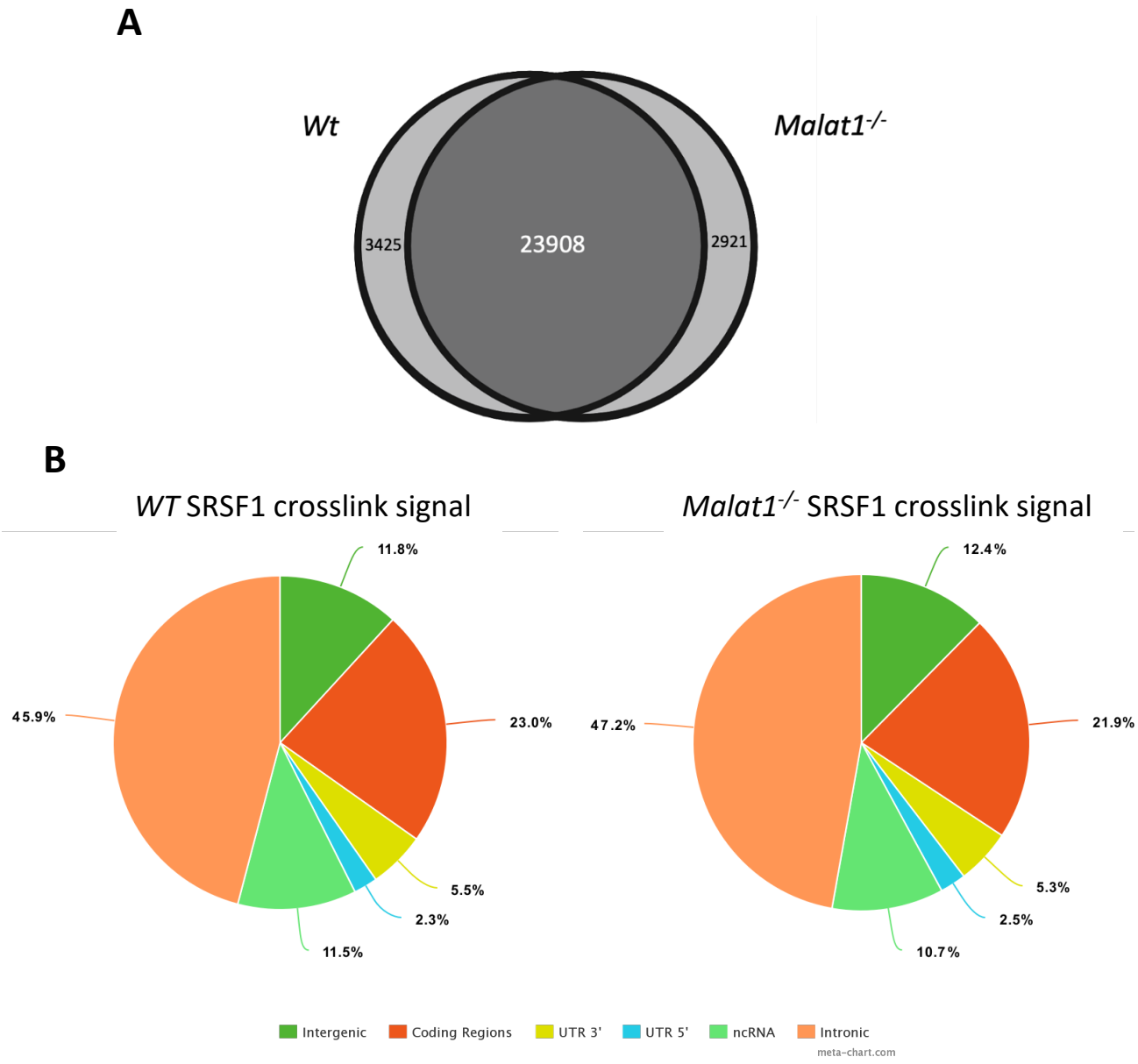


*Figure 2. Detection of SRSF1 enriched RNA transcripts in Th2 cells by iCLIP*

(A) Venn Diagram showing the number of transcripts displaying SRSF1 crosslink events detected by the iCLIP process in *WT* and *Malat1*$^{-/-}$ Th2 cells. (Generated using Venny 2.1.0) (B) Pie chart showing the proportion of SRSF1 crosslink events mapped onto 3' or 5' UTRs, coding regions (collectively making up exonic sequences), introns, intergenic and non protein coding RNAs, which were assessed by iCLIP. (Generated on Metachart.com)

cell differentiation, iCLIP was performed to gain a quantitative measure of SRSF1-RNA crosslink events in late-stage *WT* and *Malat1*$^{-/-}$ Th2 cells. The process of iCLIP in *WT* and *Malat1*$^{-/-}$ Th2 cells provided a summary of bound RNA transcripts to SRSF1 as well as a quantified value for SRSF1 crosslink events (Figure 3) and single nucleotide resolution of crosslink sites within each detected RNA transcript (Hupertz, et al., 2014).

### 3.1.1 General characteristics of SRSF1 crosslink events within Th2 cell iCLIP dataset

Sequencing of RNA transcripts crosslinked with SRSF1 provided crosslink events for 27,333 different transcripts in *WT* Th2 cells and 26,829 in *Malat1*$^{-/-}$ Th2 cells. 23,908 transcripts were common across both conditions (Figure 2a). Compared to the 4858 transcripts with recorded SRSF1 binding sites published on the official CLIP database (CLIPdb) (Yang, et al., 2015), 4764 (98%) of these transcripts were found within the total 30923 transcripts within our Th2 iCLIP dataset. The SRSF1 crosslink events of each transcript will hereafter be referred to as SRSF1 crosslink signal (normalised value of SRSF1 crosslink events per million recorded).

Concerning mapping of SRSF1 crosslink events upon the detected transcripts, we found that close to half of events were mapped onto intronic sequences across *Malat1*$^{-/-}$ and *WT* conditions (47.2%, 45.8%). This was greater than exonic sequences which contained ~30% of all SRSF1 crosslink events were mapped to in both Th2 conditions. 11.8% and 12.4% respective to *WT* and *Malat1*$^{-/-}$ Th2 cells of total SRSF1 crosslink events were mapped to uncharacterised loci and therefore will not be considered within this analysis (Figure 2b). Two replicates for each *WT* and *Malat1*$^{-/-}$ Th2 cells were analysed, with high correlation in SRSF1 crosslink signal and were pooled together for further analysis (Figure 3).

### 3.1.2 Distribution of SRSF1 binding across the transcriptome.

We wanted to discern the distribution of SRSF1 crosslink signal across all detected transcripts within the iCLIP datasets. Within the distribution of total number of transcripts for each given SRSF1 crosslink signal value, the majority of transcripts have a relatively low SRSF1 crosslink signal in both *WT* and *Malat1*$^{-/-}$ conditions. This is represented by a skewing of the distribution of transcripts towards a lower SRSF1 crosslinks signal (Figure 4a). The cumulative distribution of number of transcripts of a given SRSF1 crosslink signal shows that 95% of transcripts have an SRSF1 crosslink signal lower than 143 (*WT*) or 147 (*Malat1*$^{-/-}$) (Figure 4b). These values represent just ~0.5% of that of the transcript with the highest SRSF1 crosslink signal, *Malat1* (30,399 Crosslink Events per million). These distributions together represent a large skewing

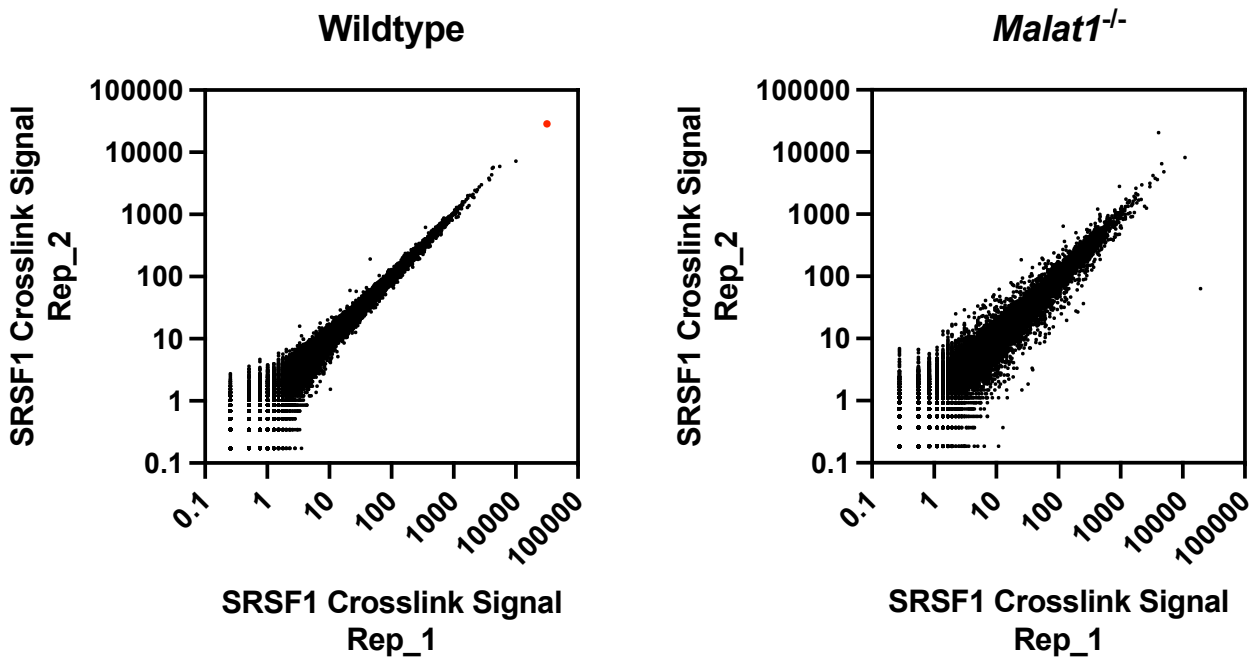


*Figure 3. Two iCLIP replicates show high correlation of SRSF1 crosslink signal in WT and Malat1$^{-/-}$ Th2 cells.*

Spearman correlation plot between SRSF1 iCLIP-seq replicates in Th2 cells. (Wt - Spearmans R = 0.9114) (*Malat1*$^{-/-}$ Spearmans R = 0.9045). Positive correlation is significant in both plots ($p < 0.001$). Red point within *WT* plot represents the SRSF1 crosslink signal of *Malat1*.

of the distribution of transcripts towards the low end of SRSF1 crosslink events, which is indicative of noise within the iCLIP data (Chen, et al., 2020). Further evidence for this comes when considering how the total SRSF1 crosslink events are distributed through the transcripts within the datasets. Considering the lower 95% percent of transcripts in terms of SRSF1 crosslink signal, this group contains just ~35% (*WT*) and ~38% (*Malat1*$^{-/-}$) of total SRSF1 Crosslink Events (Figure 4c). This shows that the noise, whilst representing the majority of detected transcripts, contains a disproportionately small amount of SRSF1 crosslink events. Considering that ~11% of total crosslink events are in intergenic/unmapped genomic regions, the 95$^{th}$ percentile (top 5%) of transcripts contained a disproportionally high 50%-53% of total SRSF1 crosslink events.

Further analysis of iCLIP dataset only considered transcripts within the 95th percentile of SRSF1 crosslink signal within either WT of *Malat1* Th2 cells. This represents a set of 1424 transcripts that were taken for further analysis.

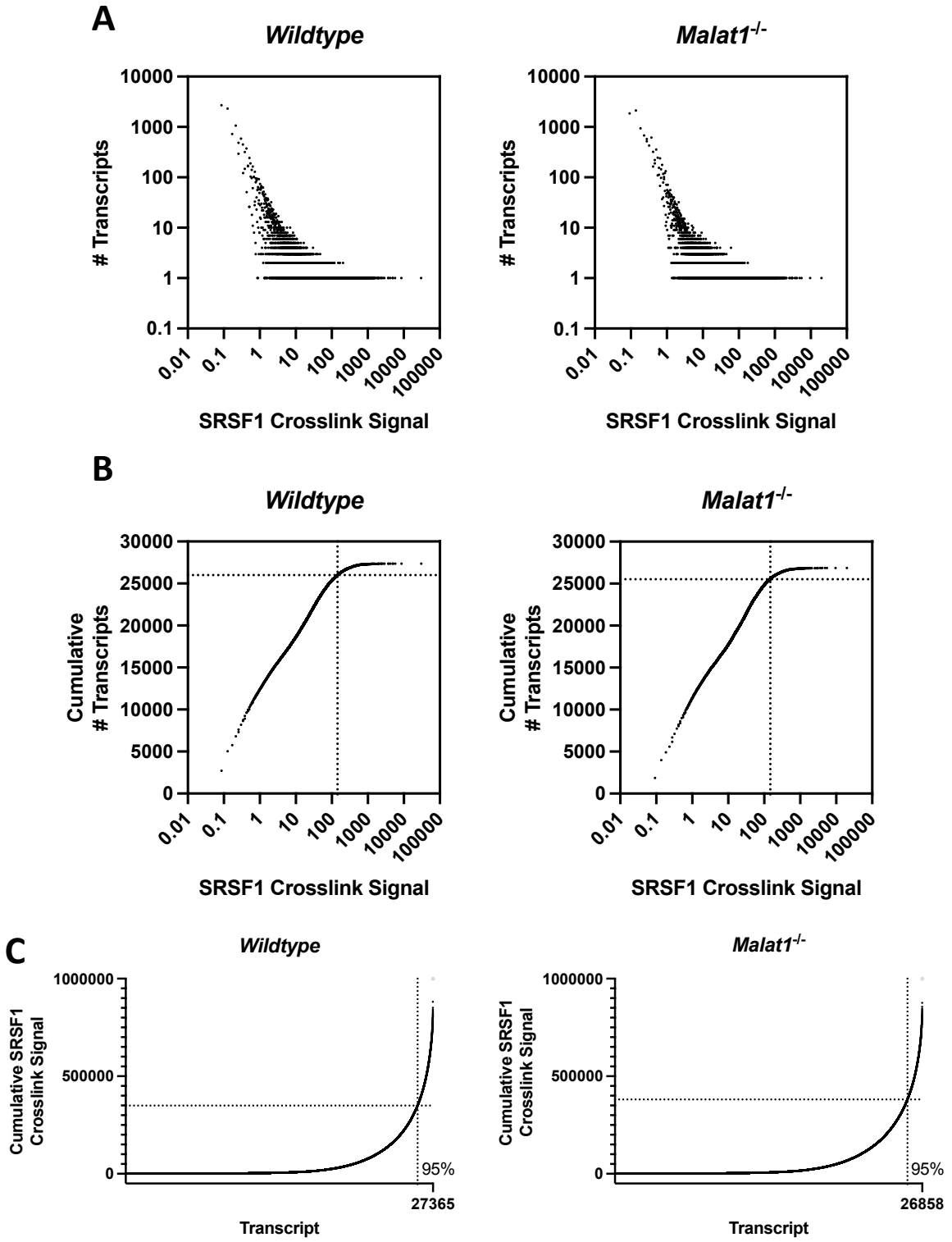


*Figure 4. iCLIP data representing the RNA binding of SRSF1 shows a majority of transcripts with relatively low SRSF1 crosslink signal*

A) Distribution and (B) cumulative distribution of SRSF1 bound transcripts in Th2 cells of each given SRSF1 crosslink signal value. (C) Cumulative Distribution of SRSF1 Crosslink signal (up to $1x10^6$) against transcripts in ascending order of SRSF1 crosslink signal. Grey points represent all SRSF1 crosslink events that are unmapped/intergenic and not considered within further analysis.

### 3.1.3 SRSF1 binding correlates with endogenous transcript abundance

To investigate how endogenous expression of transcripts influences their SRSF1 crosslink signal we used an independent RNAseq dataset of 6-day-post-activation Th2 cells. We found that there was a positive significant correlation between SRSF1 crosslink signal and transcript counts per million reads mapped (CPM) in both *WT* and *Malat1*$^{-/-}$ Th2 cells (Figure 5a,b). This effect was seen both within the whole dataset and with the established 95$^{th}$ percentile threshold applied. However, the Spearmans R value is much lower within the 95$^{th}$ percentile compared to the whole dataset. The significant positive correlation between SRSF1 crosslink signal and expression suggests the higher a transcript is expressed the more crosslink events

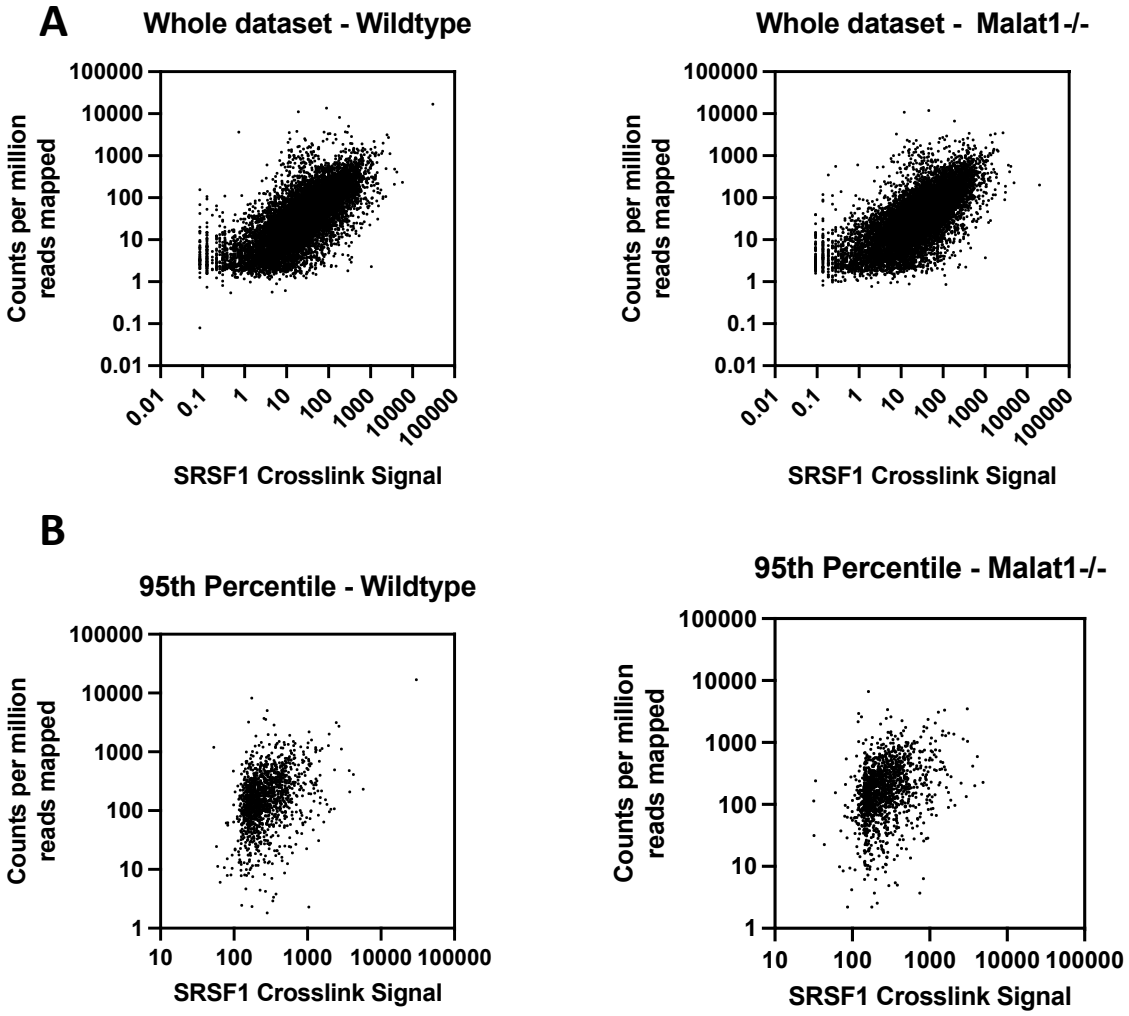


*Figure 5. Significant positive correlation is shared between transcript expression and transcript SRSF1 crosslink signal*

Spearmans correlation analysis between transcript abundance (from Th2 RNA-seq dataset) and SRSF1 crosslink signal (from iCLIP dataset) in (A) whole iCLIP dataset dataset (WT, R = 0.5523) (*Malat1*$^{-/-}$, R = 0.7196) and (B) applied 95$^{th}$ percentile threshold (WT, R = 0.3593) (*Malat1*$^{-/-}$, R = 0.3511). All plots show significant positive correlation ($p < 0.001$).

that occur between SRSF1. However, outliers of this correlation as well as the Spearmans R values being significantly less than 1 showed that there were transcripts demonstrating enrichment in SRSF1 binding that is alternative to what is expected based on their abundance.

### 3.1.4 Greatest SRSF1 crosslink signal transcripts are significantly enriched in T cell function

We performed analysis of the gene ontology (GO) in terms of biological processes (BP) for the highest transcripts in terms of SRSF1 crosslink events in *wildtype* Th2 to assess the function and general binding behaviour of SRSF1. Within both of these conditions the top 50 (SRSF1 crosslink signal > 1109) and top 500 (SRSF1 crosslink signal > 308) showed significant enrichment in gene sets involved in T cell activation and differentiation (Figure 6). This suggests an important role of SRSF1 binding to important T cell regulators in conferring its control of Th2 differentiation.

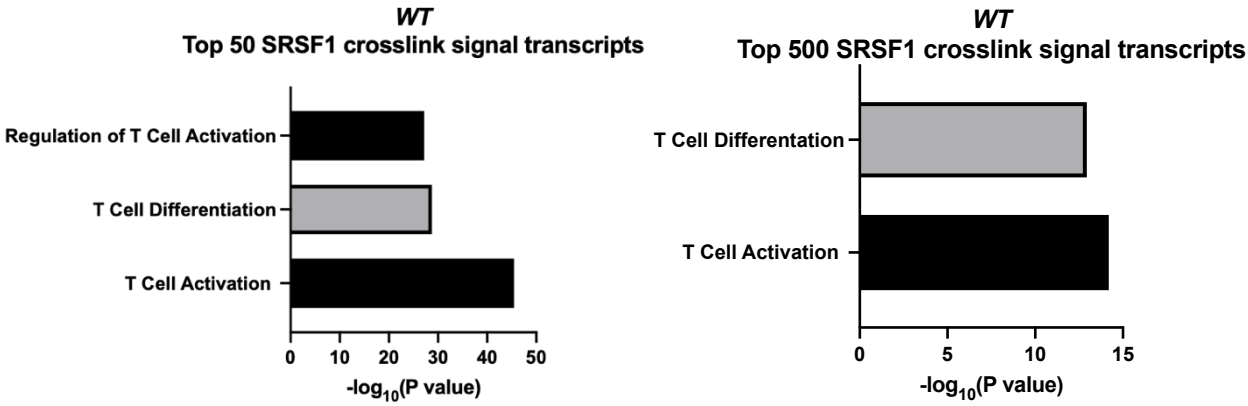


*Figure 6. Transcripts with the greatest SRSF1 crosslink signal in WT Th2 cells are significantly enriched in T cell related gene sets.* Gene Ontology (Biological Processes) enrichment of transcripts with greatest 50 and 500 SRSF1 crosslink signal in *WT* Th2 cells.

### 3.1.5 *Malat1* regulates SRSF1 RNA binding behaviour in Th2 cells

To investigate whether *Malat1* regulates SRSF1 RNA binding behaviour it is important to consider how the SRSF1 crosslink signal of transcripts changes from *WT* to *Malat1*$^{-/-}$ Th2 conditions. Using the RNA-seq dataset we determined that SRSF1 abundance was not significantly changed by the loss of *Malat1* in Th2 cells ($Log_2$(abundance fold change) = -0.07, p = 0.54). However, SRSF1 abundance at the protein level was not assessed. The SRSF1 crosslink signal of each transcript within *WT* and *Malat1*$^{-/-}$ Th2 cells shows a strong positive correlation (Figure 7a). Additionally, proportions of SRSF1 crosslink signal mapping within intronic, exonic and intergenic sequences between *Malat1*$^{-/-}$ and WT Th2 remain largely

unchanged (Figure 2a). This data suggests that *Malat1* loss does not have any effect on general SRSF1 binding behaviour to all SRSF1 enriched transcripts in *WT* Th2 cells.

The effect of individual transcript SRSF1 crosslink signal can be quantified with the value $log_2$(SRSF1 Crosslink Signal Fold change) (*$Log_2$(csFC)*). We found that 44 transcripts had a $Log_2$(csFC) > |1|, indicative of a two-fold increase or decrease in SRSF1 crosslink signal as *Malat1* is lost. Of transcripts with a $Log_2$(csFC) > |0.6|, indicative of a 1.5-fold increase or decrease in SRSF1 crosslink fold change as *Malat1* is lost, 133 were characterised. GO (BP) analysis of these transcripts reveal that whilst transcripts with $Log_2$(csFC) > |1| were not significantly enriched in any T cell related gene sets, transcripts with $Log_2$(csFC) > |0.6| were significantly enriched in gene sets relating to T cell activation, differentiation and proliferation (Figure 7b). This suggests that some of the *Malat1* mediated binding RNA transcript partners of SRSF1 are involved in CD4+ T cell activation. Of note there are several important regulators of CD4+ T cells which have a 1.5-fold or greater change in SRSF1 crosslink signal when *Malat1* is lost in Th2 cells. These include, *Gata3*, *Runx3*, *Il2ra*, *Btla*, *Ctla4* and *Ccr2*.

As previously stated, expression of a transcript is positively correlated with SRSF1 crosslink signal. Therefore, as an additional investigation, SRSF1 crosslink signal values were normalised to the CPM values of the same transcript. 55 transcripts had a normalised $Log_2$(csFC) > |1| and 174 transcripts had a normalised $Log_2$(csFC) > |0.6|. Both thresholds of normalised $Log_2$(csFC) were significantly enriched in T cell activation and differentiation gene sets (Figure 7c). Also following normalisation, some of the known CD4+ T cell regulators remained with a $log_2$(csFC) value of greater than |1|, including *Gata3*, *Btla*, *Ccr2* and *Runx3*.

Overall, this bioinformatic analysis revealed a catalogue of 232 transcripts which are differentially bound to SRSF1 by a quantified factor of 1.5-fold or more in *Malat1*$^{-/-}$ Th2 cells compared to *WT.* Additionally, several of these transcripts are well-known and characterised regulators of Th differentiation. Full details of these transcripts are available in Appendix 1. Of these transcripts, 99 were identified exclusively through normalised of $Log_2$(csFC). 58 transcripts, whilst displaying $Log_2$(csFC) > |0.6| when not considering abundance, did not show this magnitude of fold change upon normalisation for transcript abudance. 75 transcripts displayed $Log_2$(csFC) > |0.6| both when not considering abundance and when normalisation was carried out. Due to the function of SRSF1 as a post-transcriptional regulator being linked to direct binding of RNA targets, we wanted to investigate if these candidate

transcripts may also be differentially regulated as well as differentially bound by SRSF1 upon *Malat1* loss.

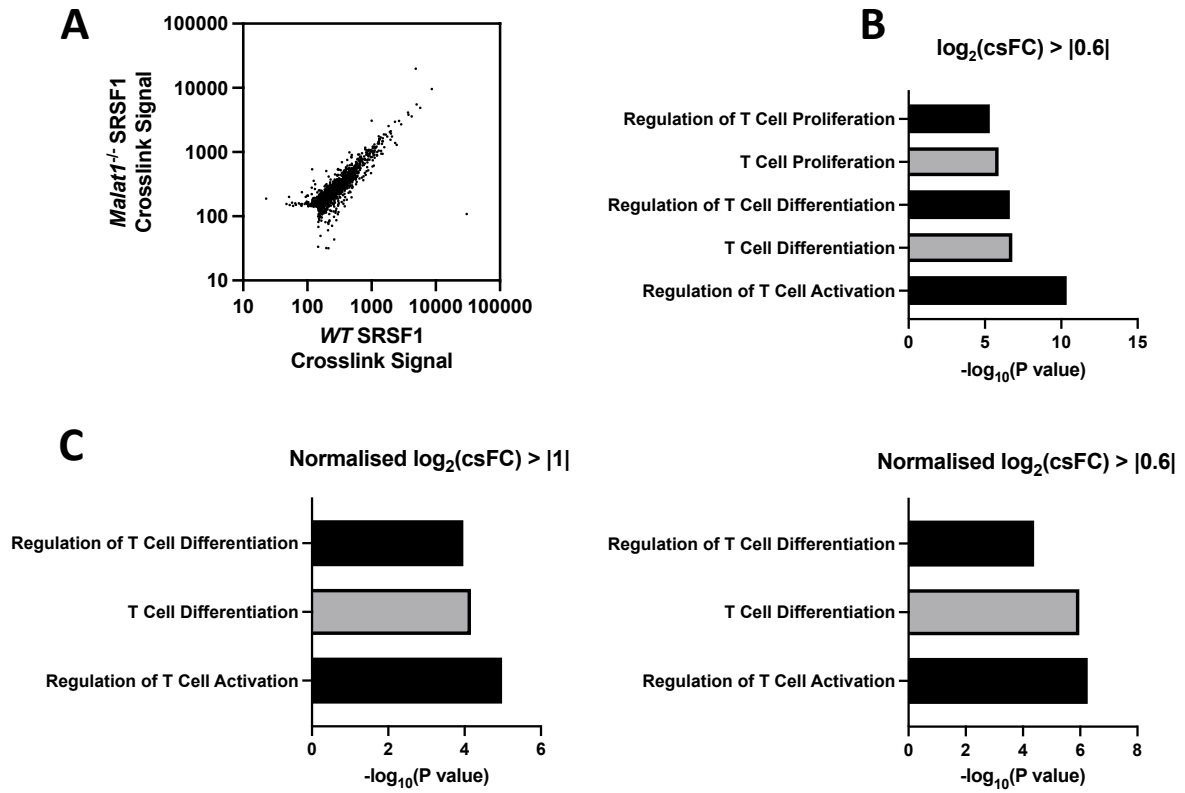


*Figure 7. Malat1 regulates SRSF1 binding to transcripts significantly enriched in T cell related gene sets.*

(A) Correlation of SRSF1 Crosslink Events of transcripts within the 95$^{th}$ percentile of WT and *Malat1*$^{-/-}$ Th2 cell iCLIP datasets (Spearmans R = 0.8593, p < 0.001). *Gene ontology (Biological processes)* enrichment graphs of transcripts with (B) log2(csFC) > |0.6| and (C) normalised log2(csFC) > |1| or |0.6|.

## 3.2 Assessment of transcriptional and post-transcriptional regulation of *Il2ra* and *Runx3* by *Malat1* in Th2 cells.

Leading on from the previous bioinformatic analysis we were keen to investigate the effect of *Malat1* loss on CD4+ T cell relevant transcripts with differential binding to the SRSF1. We specifically investigated changes in abundance and specific isoform usage of candidate transcripts within *in vitro* activated Th2 cells.

### 3.2.1 *In vitro* Th2 differentiation of naïve CD4+ T cells

*WT* and *Malat1*$^{-/-}$ naïve CD4+ T cells differentiated *in vitro* into Th2 cells *(carried out by Magnus Gywnne)*. Flow cytometry demonstrated a significant increase of the Th2 activation markers and hallmark cytokines IL4 and IL10 from 4 to 6 days post-activation (Figure 8). These activated Th2 cells were used to investigate the isoform usage and abundance of candidate transcripts identified through SRSF1 iCLIP upon *Malat1* loss.

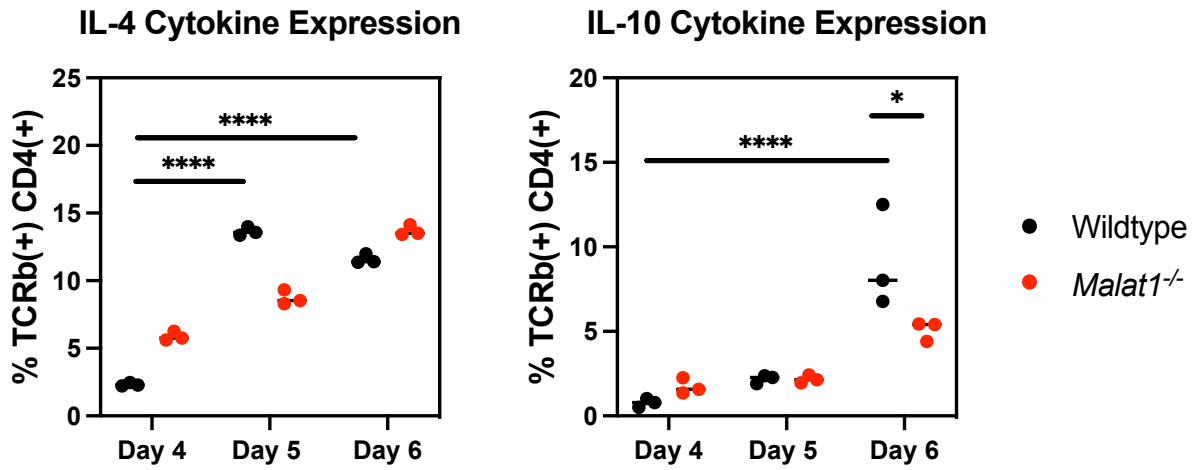


*Figure 8. rIL2 treatment of CD4+ T cells causes a significant increase in IL4 and IL10 expressing cells.*

4 days post anti-CD3+anti-CD28 treatment CD4+ T cells received rIL-2 treatment. Following this flow cytometry was used to analyse the percentage of IL-4 and IL-10 positive cells at 24 hour intervals, 4-6 days post anti-CD3 treatment. One-way anova and subsequent multiple comparison test determined statistical significance of results. Concerning P values: **** < 0.0001, * < 0.05.

### 3.2.2 *Malat1* loss inhibits SRSF1 binding to *Il2ra* but does not affect *Il2ra* abundance or transcript usage

*Il2ra* SRSF1 crosslink signal decreases in Th2 cells as *Malat1* is lost ($log_2(CSFC)$ = -0.87). IL2RA is a cell surface receptor, which along with IL2RB and IL2RG make up the IL2 receptor found on CD4+ T cells (Minami, et al., 1993). IL2 is an essential cytokine during CD4+ T cell activation and stimulates cell proliferation and expression of effector cytokines (Ross & Cantrell, 2018). *Il2ra* pre-mRNA can undergo an alternative splicing event, in which the pre-mRNA is truncated short, losing exons 2-8 (Figure 9a). Mapping the SRSF1 crosslink events across the *Il2ra* transcript shows a peak in SRSF1 binding at the alternative splicing site that produces the truncated *Il2ra* isoform. This is also the region of the *Il2ra* truncated isoform 3' UTR (Figure 9b). Within *Malat1*$^{-/-}$ Th2 cells SRSF1 binding at the same position of the *Il2ra* transcript is altered to a disproportionate extent at two nucleotide locations (Figure 9c). An altered binding preference for *Il2ra* at this alternative splice site by SRSF1 when *Malat1* is lost

suggests that *Malat1* loss is indirectly regulating isoform usage of long *Il2ra* and truncated *Il2ra* mRNA.

To investigate this *Il2ra* expression was assessed across *in vitro* activated Th2 cells. *Il2ra* levels were investigated at day 4 post activation and onwards, due to this timepoint being when Th2 cells receive exogenous IL2 *in vitro*. Three sets of primers were designed, one to amplify a control region specific to all *Il2ra* isoforms and two to each specifically amplify unique regions within either long or truncated *Il2ra* (Figure 10a). These primers sets were then validated in

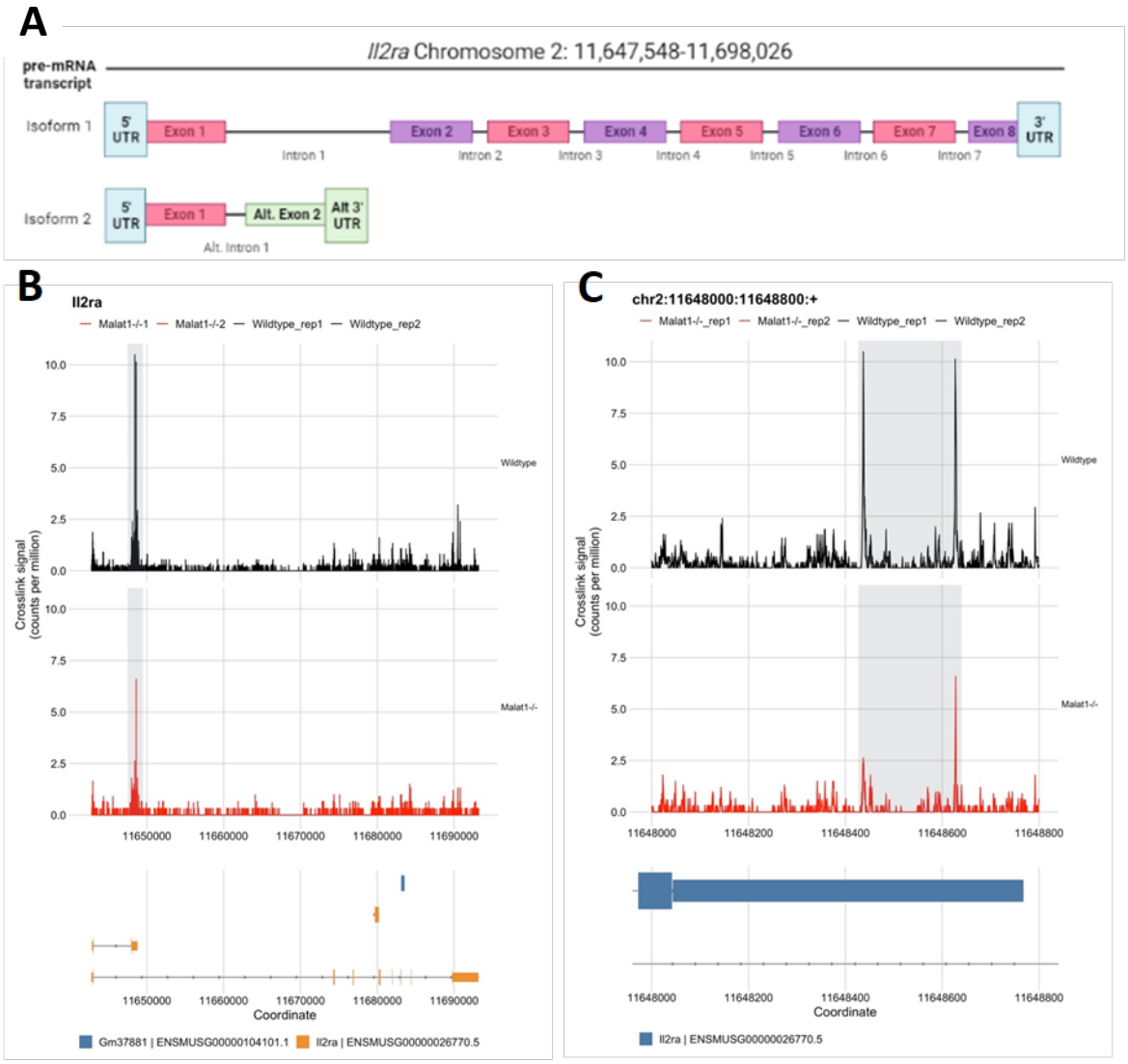


*Figure 9. Malat1 regulates SRSF1 binding with Il2ra specifically at the truncated Il2ra splicing site.*

(A) Schematic representation of long *Il2ra* pre-mRNA and truncated *Il2ra* pre-mRNA. CLIP-plots showing distribution of SRSF1 crosslink events across (B) the whole *Il2ra* pre-mRNA transcript and (C) the 5' UTR of truncated *Il2ra*. Highlighted grey portions show SRSF1 enriched portion of truncated *Il2ra* 3'UTR.

serial dilutions of cDNA to determine amplification efficiency (accepted between 90-110%, Figure 10b) as well as ensuring a single amplicon product (Figure 10c).

RT-qPCR analysis with the differentiated Th2 cells showed that *Malat1* has no significant effect on *Il2ra* expression or isoform usage at the time points investigated (Figure 11a-c).

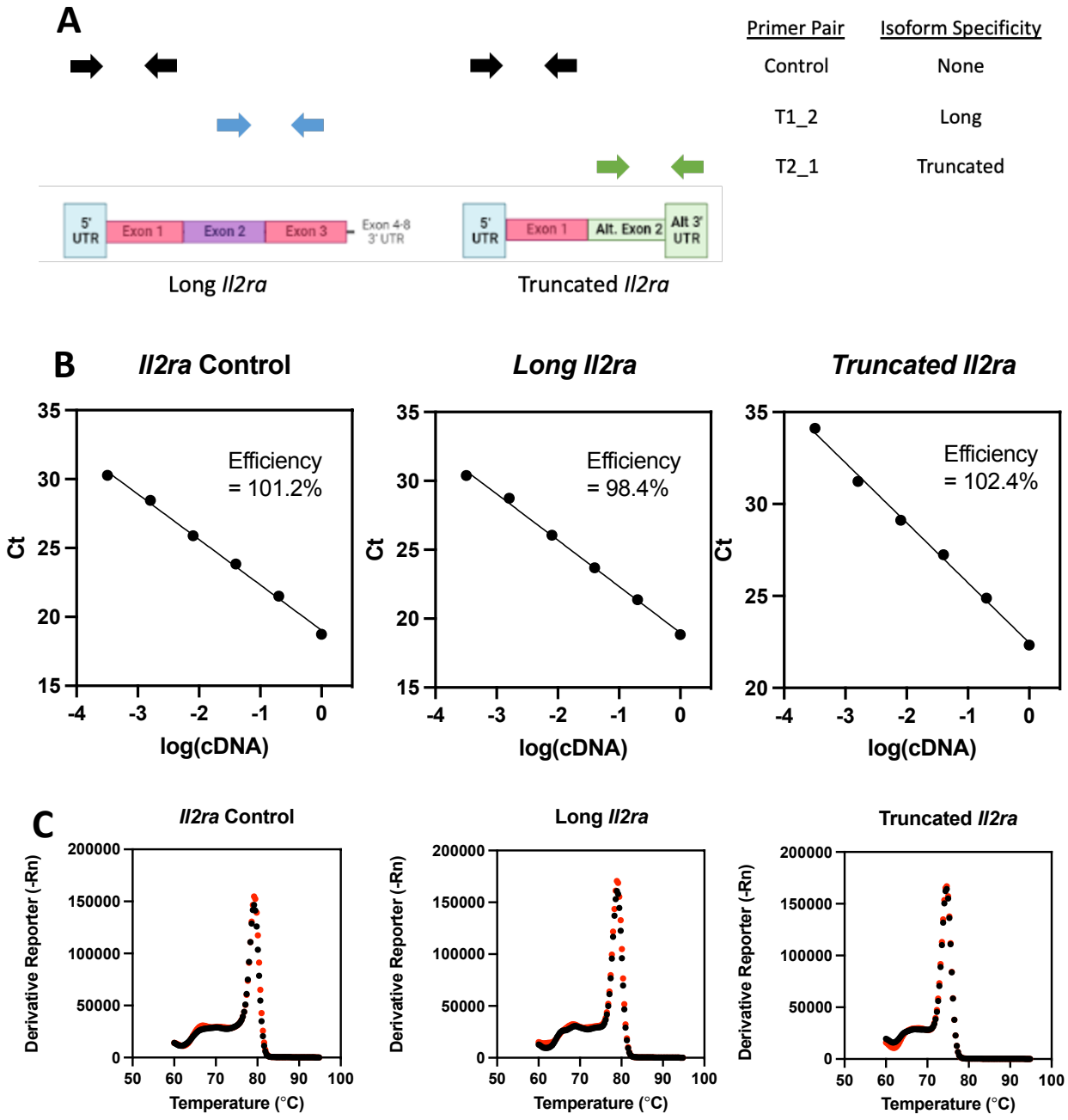


*Figure 10. RT-qPCR primers sets were designed to assess Il2ra abundance and isoform usage.*
(A) Schematic of the *Il2ra* mRNA structure showing the alternatively spliced truncated *Il2ra* isoform. Different primer pairs and their location relative to the *Il2ra* mRNA isoforms are shown as arrows. (B) cDNA amplification efficiency of *Il2ra* primer pairs determined by gradient of RT-qPCR Ct value output across serially diluted cDNA samples. (C) Melt curves of *Il2ra* primers pairs, shown as increasing temperature of RT-qPCR reaction against the negative gradient of SYBR Green fluorescence. Peaks are indicating amplification products.

Additionally, *Malat1* does not significantly change the ratio at which the *Il2ra* isoforms are expressed across Th2 activation (Figure 11d). Of note the long *Il2ra* isoform is expressed at much higher levels (10-20-fold), showing that the truncated *Il2ra* isoform does not have prevalent expression within Th2 cells compared to long *Il2ra*. These results indicate that although SRSF1 binding to *Il2ra* mRNA is altered upon *Malat1* loss, this neither affects *Il2ra* expression nor isoform usage in Th2 cells.

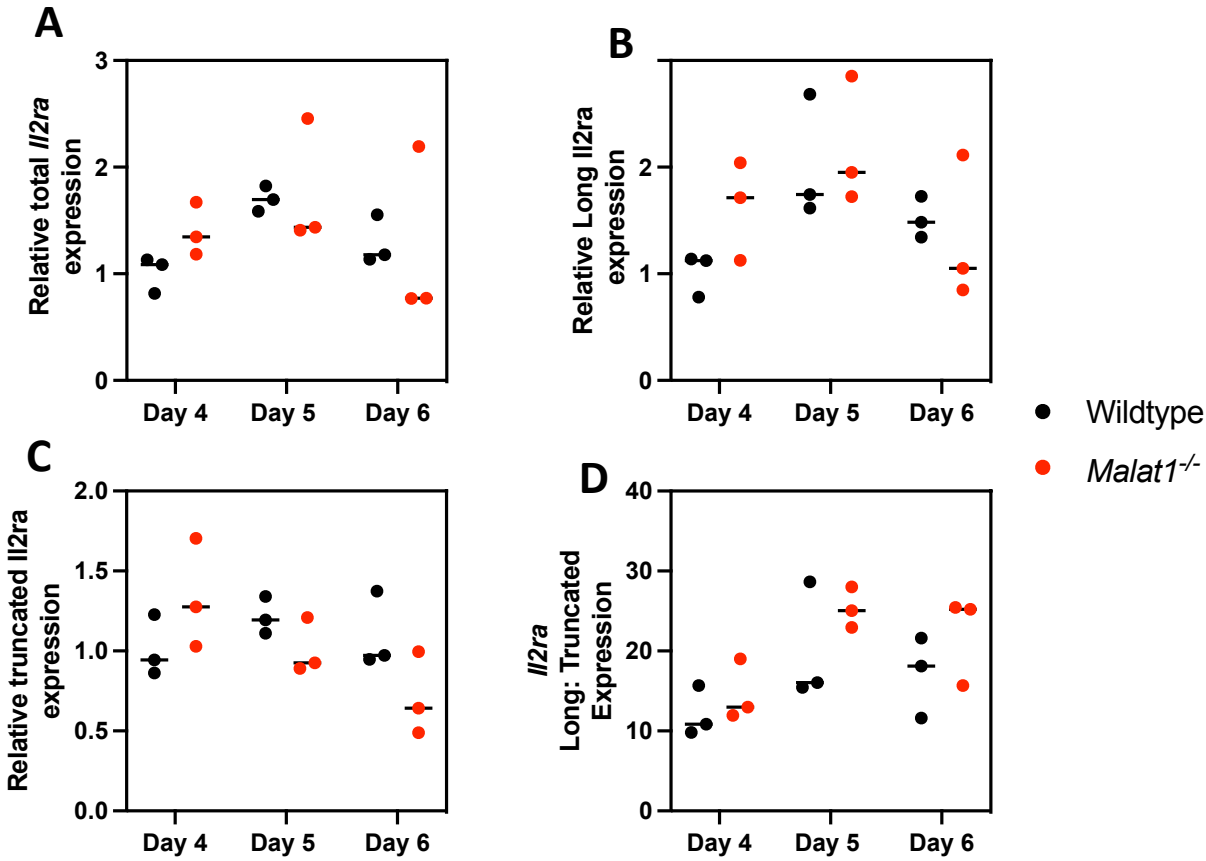


*Figure 11. Malat1 does not regulate Il2ra expression or alternative splicing in Th2 cells.*

Relative mRNA levels of *(A)* total *Il2ra, (B) long Il2ra* and (C) truncated *Il2ra* from 4-6 days post activation of CD4+ Naïve cells. (D) Ratio of long *Il2ra* against truncated *Il2ra* in terms of relative mRNA abundance. All mRNA abundance levels are normalised to U6 expression and relative to *Il2ra* expression in at day 4 post activation. One-way anova and subsequent multiple comparison test determined statistical significance of results. No significance between *WT* and *Malat1*$^{-/-}$ expression/ratio values was viewed at each given timepoint between *WT* and *Malat1*$^{-/-}$ Th2 cells.

### 3.2.3 *Malat1* loss increases SRSF1 binding to *Runx3* whilst reducing *Runx3* abundance and altering isoform usage

*Runx3,* a vital regulator in maintaining naïve CD4+ T cell differentiation towards proper Th2 subtype (Kohu, et al., 2009), displays increased binding to SRSF1 as *Malat1* is lost in Th2 cells ($log_2$(csFC) = 0.66). *Malat1* loss has the opposite effect on *Runx3* expression in Th2 cell, with it being downregulated at the mRNA level ($log_2$(abundance fold change) = -0.77). A positive $log_2$(csFC) and a decrease in abundance upon *Malat1* loss, therefore, lead to *Runx3* having a relatively large normalised $log_2$(csFC) of 1.43 (2.7-fold increase). This indicates with greater confidence that *Malat1* has a direct regulatory role in SRSF1 binding to *Runx3*. *Runx3* can be transcribed via two separate promotors, distal and proximal (Figure 12a). *Runx3* transcription via the proximal promotor (*pRunx3*) produces a transcript with poorer translation efficiency compared to *Runx3* transcripts derived from the distal promotor (*dRunx3*) (Kim, et al., 2015). Whilst not displaying any notable or disproportionate peaks, analysis of the mapping of SRSF1 crosslink events across *Runx3* showed that there was a cluster of peaks specific to distal *Runx3* (*dRunx3*) (Figure 12b). SRSF1 crosslink events are greater at the 5'UTR of *dRunx3* over that of *pRunx3*. Considering *Malat1*$^{-/-}$ Th2 cells, SRSF1 binding to *Runx3* is increased, notably in the *dRunx3* specific 5'UTR and portion of the 3'UTR (Figure 12c,d,e). This mapping data is indicative of preferential binding by SRSF1 to *dRunx3*, which is increased when *Malat1* is lost.

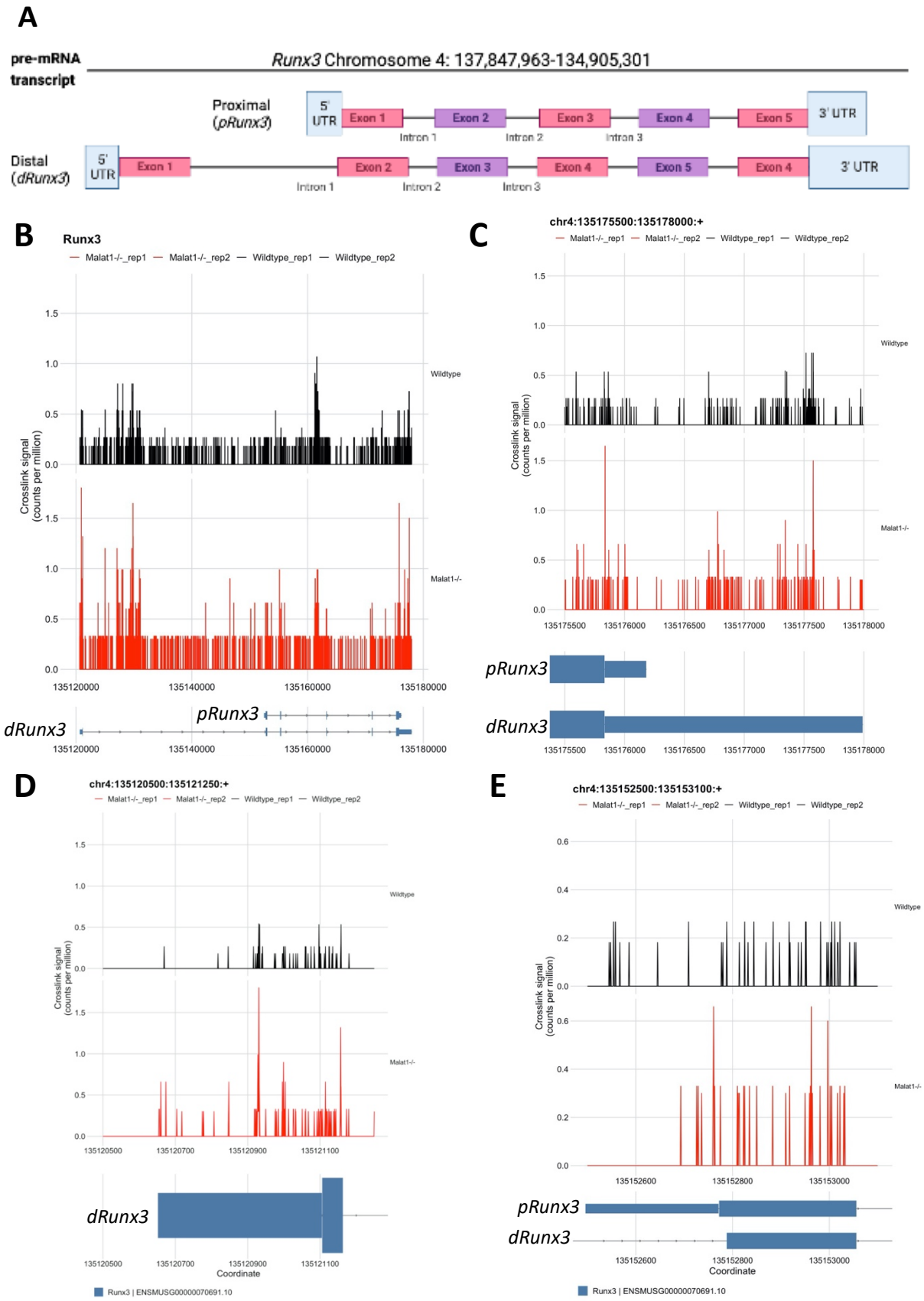


*Figure 12. Malat1 regulates SRSF1 binding to the distal isoform of Runx3.*

(A) Schematic representation of transcripts expressed from alternative promotors from the *Runx3* gene locus, *dRunx3* and *pRunx3*. The distribution of SRSF1 crosslink events across the (B) whole *Runx3* pre-mRNA transcript, (C) *dRunx3* and *pRunx3* 3' UTR, (D) *dRunx3* 5' UTR and (E) *pRunx3* 5' UTR.

To assess the effect of *Malat1* on *Runx3* abundance of promotor specific isoform usage in Th2 cells, three sets of primers were designed. One to amplify a control region specific to both *Runx3* isoforms and two to each specifically amplify unique regions within either distal or proximal *Runx3* (Figure 13a). Primers sets were then validated in serial dilutions of cDNA to determine amplification efficiency (accepted between 90-110%, Figure 13b) and to ensure a single amplicon product (Figure 13c).

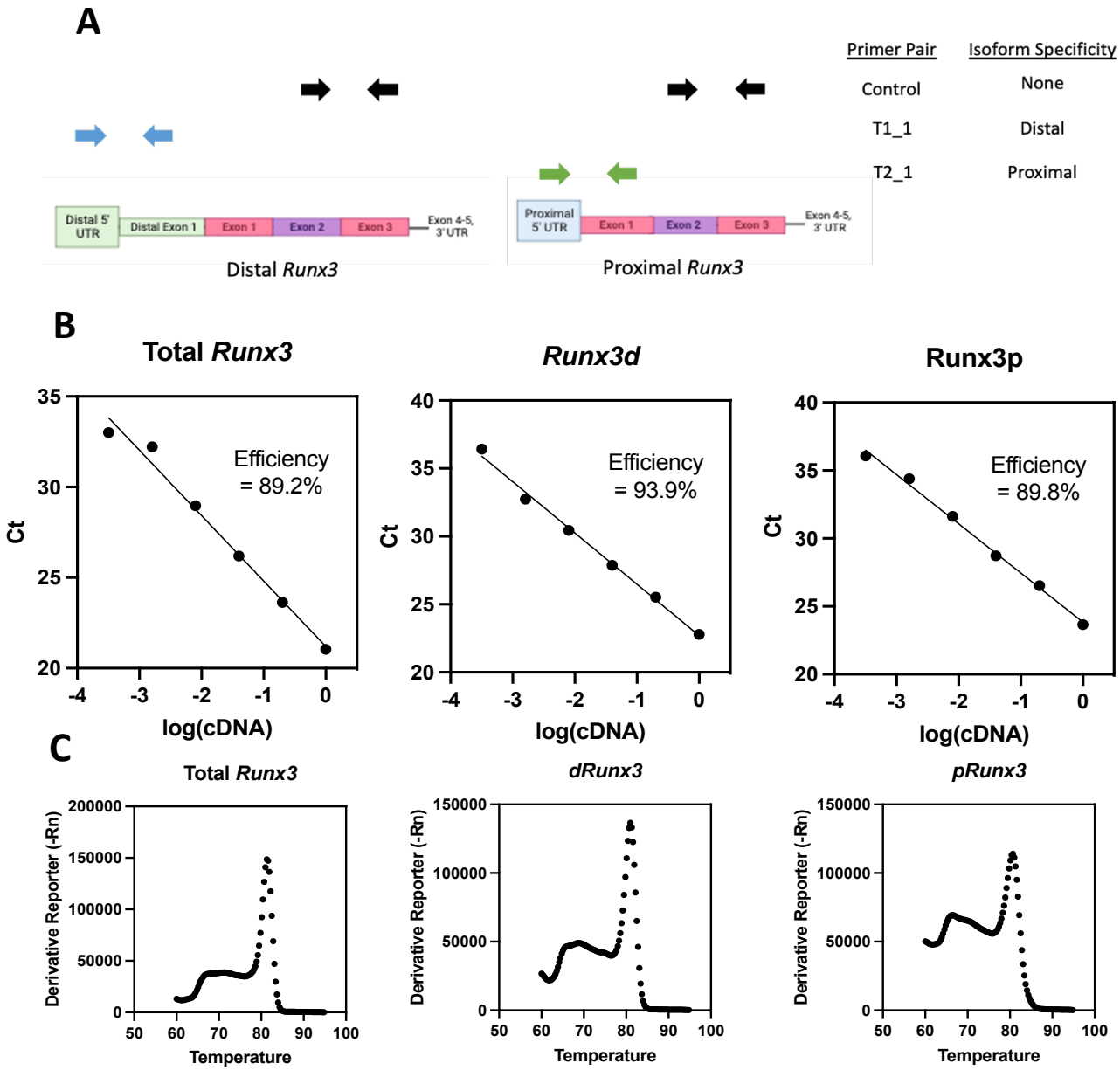


*Figure 13. RT-qPCR primers sets were designed to assess Runx3 abundance and isoform usage.*

*(A)* Schem*atic* of the *Runx3* mRNA structure *showing the alternative promotor transcribed products of the Runx3 gene.* Different primer pairs and their location relative to the *Runx3* mRNA *isoforms* are shown as arrows. *(B) cDNA ampliciation efficiency of Runx3 primer pairs determined by gradient of RT-qPCR Ct value output across serially diluted cDNA samples. (C) Melt curves of* Runx3 *primers pairs, shown as increasing temperature of RT-qPCR reaction against the negative gradient of SYBR Green fluorescence.* Peaks are indicating amplification products.

RT-qPCR experiments included 4 to 6-day post activation timepoints along with additional earlier timepoints taken from an independent CD4+ *in vitro* activation experiment. This was due to *Runx3* action as an important transcriptional regulator, which is important as the large transcriptional shift occur immediately following activation of CD4+ T cells (Ip, et al., 2007). Early time points (both 0 and 24hr post activation), show that whilst *Malat1* loss causes significantly lower levels of *pRunx3 prior to* Th2 specific induction signals of CD4+ T cells, total *Runx3* or *dRunx3* abundance do not change (Figure 14). At later time points of 4 days-post activation and onwards, results show significant downregulation of total *Runx3* and *dRunx3* at 6 days post activation in *Malat1*$^{-/-}$ Th2 cells, whilst *pRunx3* was not significantly changed

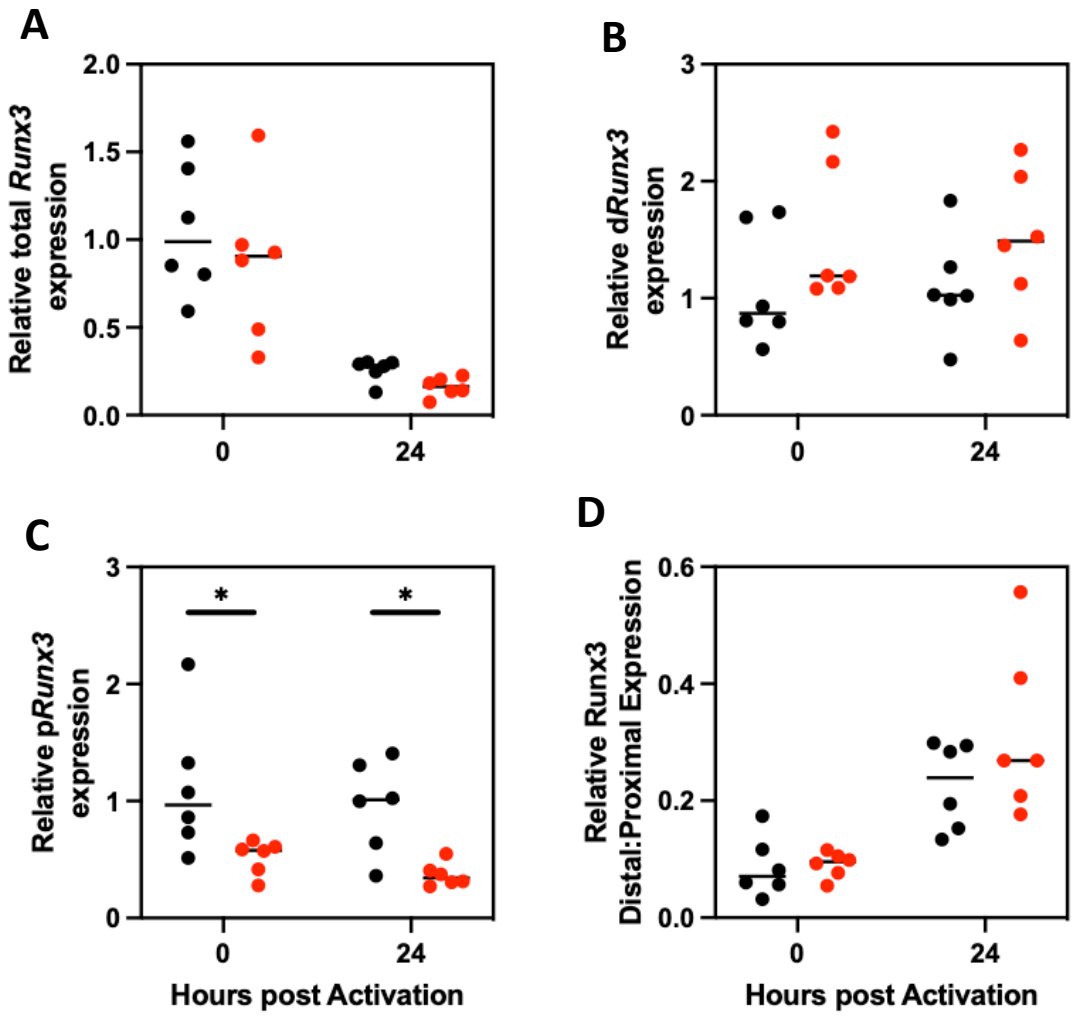


*Figure 14. Malat1 loss leads to lower pRunx3 isoform usage during Th2 activation.*

Relative mRNA levels of (A) total *Runx3,* (B) distal *Runx3* and (C) proximal *Runx3* from 0-24 hours post activation of CD4+ Naïve cells. (D) Ratio of dRunx3 against pRunx3 in terms of relative mRNA abundance. All mRNA abundance levels are normalised to U6 expression and relative to *Runx3* expression at 0hr. One-way anova and subsequent multiple comparison test determined statistical significance of results. Concerning P values: * = $p < 0.05$.

(Figure 15). These changes strongly suggest a downstream decrease in RUNX3 protein levels (Kim, et al., 2015), however this was not investigated within this study. Overall, these results show that *Malat1* loss causes both an increase in SRSF1 binding to *Runx3,* as well as causing decreased abundance and altered promotor specific *Runx3* isoform usage in Th2 cells.

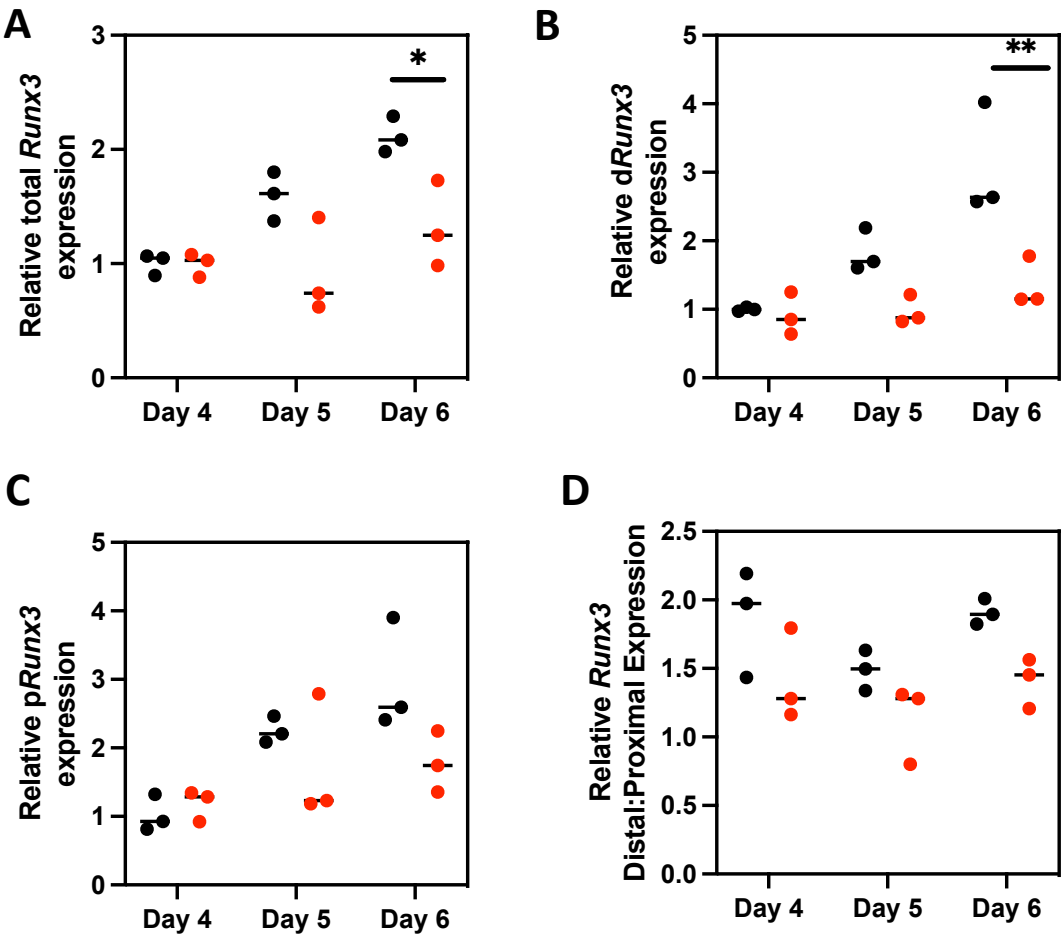


*Figure 15. Malat1 regulates total Runx3 and dRunx3 isoform expression following IL2 treatment in differentiated Th2 cells.* Relative mRNA levels of (A) total *Runx3,* (B) distal *Runx3* and (C) proximal *Runx3* from 4-6 days post activation of CD4+ Naïve cells. IL2 treatment occurred 4 days post activation and Day 4 samples were taken prior to treatment (B) Ratio of dRunx3 against pRunx3 in terms of relative mRNA abundance. All mRNA abundance levels are normalised to U6 expression and relative to Runx3 expression 4 days post activation. One-way anova and subsequent multiple comparison test determined statistical significance of results. Concerning P values: ** = $p < 0.01$, * = $p < 0.05$.

## 3.3 Production of *in vitro* CRISPR-Cas9 mediated SRSF1 knockout models

We planned to utilise the CRISPR-Cas9 system to produce *in vitro* SRSF1 knockout models to investigate the post-transcriptional regulation of candidate transcripts by SRSF1 during CD4+ T cell activation and differentiation. CRISPR-Cas9 is an accessible method of genetic editing that in our context was used to target and cleave the *Srsf1* gene to knockout functional SRSF1

protein within EL4 cell lines and naïve CD4+ T cells. Gene knockouts were achieved through transfection of CRISPR-Cas9 RNPs into cells. CRISPR RNA (crRNA) within these RNPs was used to direct Cas9 nuclease mediated double stand DNA cleavage. Electroporation was used to transiently permeabilise the cell membrane to allow RNPs to pass through.

### 3.3.1 Optimisation of Electroporation conditions

Initially, we aimed to develop and optimise a protocol for successful gene knockout in EL4 cells by first targeting HPRT as a positive control gene. Published optimal EL4 cell line electroporation conditions were not available so initial conditions were according to the IDT protocol for RNP transfection to Jurkat cells (human T cell lymphoma cell line) (Integrated DNA Technologies, 2022). Following electroporation, both the HPRT crRNA and negative control crRNA associated RNP treated EL4 cell populations experienced mass cell death (Figure 16a). However, recovery of cell populations did occur after 168 hours to a degree where sufficient cell samples could be extracted protein analysis. HPRT protein abundance

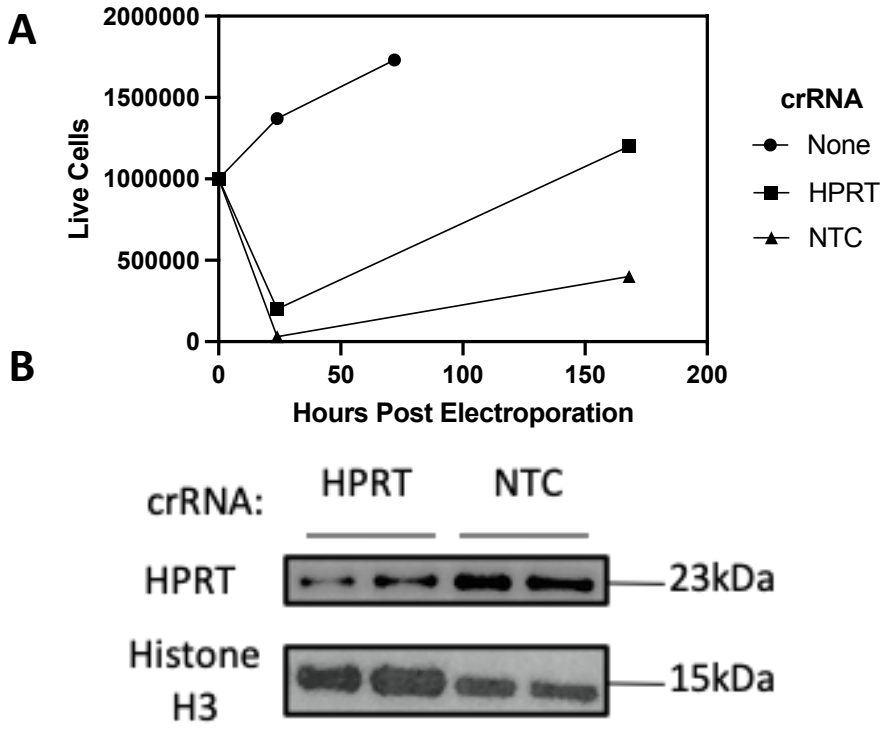


*Figure 16. Electroporation of EL4 cells with HPRT targeting crRNA associated RNPs leads to a reduction in HPRT abundance.* (A) Viability of cells following Electroporation with HPRT and negative control (NTC) crRNA associated CRISPR-Cas9 associated RNPs (1µM) at a starting cell number of $1x10^6$. Thw crRNA negative category did not receive electroporation. (B) Western blot results showing HPRT abundance in EL4 cells 168hrs post electroporation. Electroporation was carried out at 180V, 2ms, 1 pulse using Gene Pulser XCell system.

was lower in the HPRT crRNA treated EL4 cells compared to negative control crRNA treated cells (Figure 16b). However, HPRT protein expression still being clearly present in the expanded HPRT crRNA treated population suggests a low transfection and knockout efficiency. Nevertheless, these results validated this method of CRSIPR-Cas9 mediated targeted knockout and transfection using these RNP reagents could be used to reduce target protein abundance within EL4 cells.

Prior to continuing with SRSF1 knockout within EL4 cells we were keen to improve post electroporation viability. Whilst a reduction in viability is common in electroporation (Napotnik, et al., 2021), large losses of viability up to 90% is not practical and required improvement via experimentational optimisation. To optimisation electroporation conditions, EL4 cells were electroporated in mock transfection experiments without the presence of RNPs, using the Gene Pulser Xcell system at varying voltage (180V/200V/220V), pulse number (1/2) and cell density ($1x10^6$/$2x10^6$ EL4 cells per 100ul reaction). Subsequent viability of the electroporated EL4 cells were analysed showing all conditions resulted in an 80-90% drop in viability closely following electroporation compared to non-electroporated controls (Figure 17a). This was similar to the drop seen in Figure 16a, where EL4 cells were electroporated with HPRT targeting RNPs, indicating no improvement. Following these results, we opted to test a further parameter and use a larger microcuvettes for electroporation. However, similar results were gained from here, with a drastic reduction in EL4 cell viability occurring closely after electroporation at 24hrs across all conditions (Figure 17b). Subsequent to this, we carried out further optimisation through improvement of reaction conditions in EL4 transfections which involved RNPs. This included keeping the cuvettes at a cold temperature throughout the experiments and immediate transfer of electroporated cells to into cold DMEM media. From these amendments a sustainably high viability after 24 hours (~60-90%) was achieved across several electroporation conditions, including that of higher voltages (220Vs) and pulse numbers (2 pulses) (Figure 17c).

### 3.3.2 SRSF1 knockout in EL4 cells

Following optimisation of Gene Pulser Xcell system conditions we began experimentation to achieve a clonal *Srsf1*$^{-/-}$ EL4 population. Transfection was carried out with EL4 cells along with CRISPR-Cas9 RNPs containing SRSF1-AA crRNA, SRSF1-AC crRNA or a combination of both. Initial transfection at 180V and for 1 pulse did not yield any detectable loss in SRSF1 protein

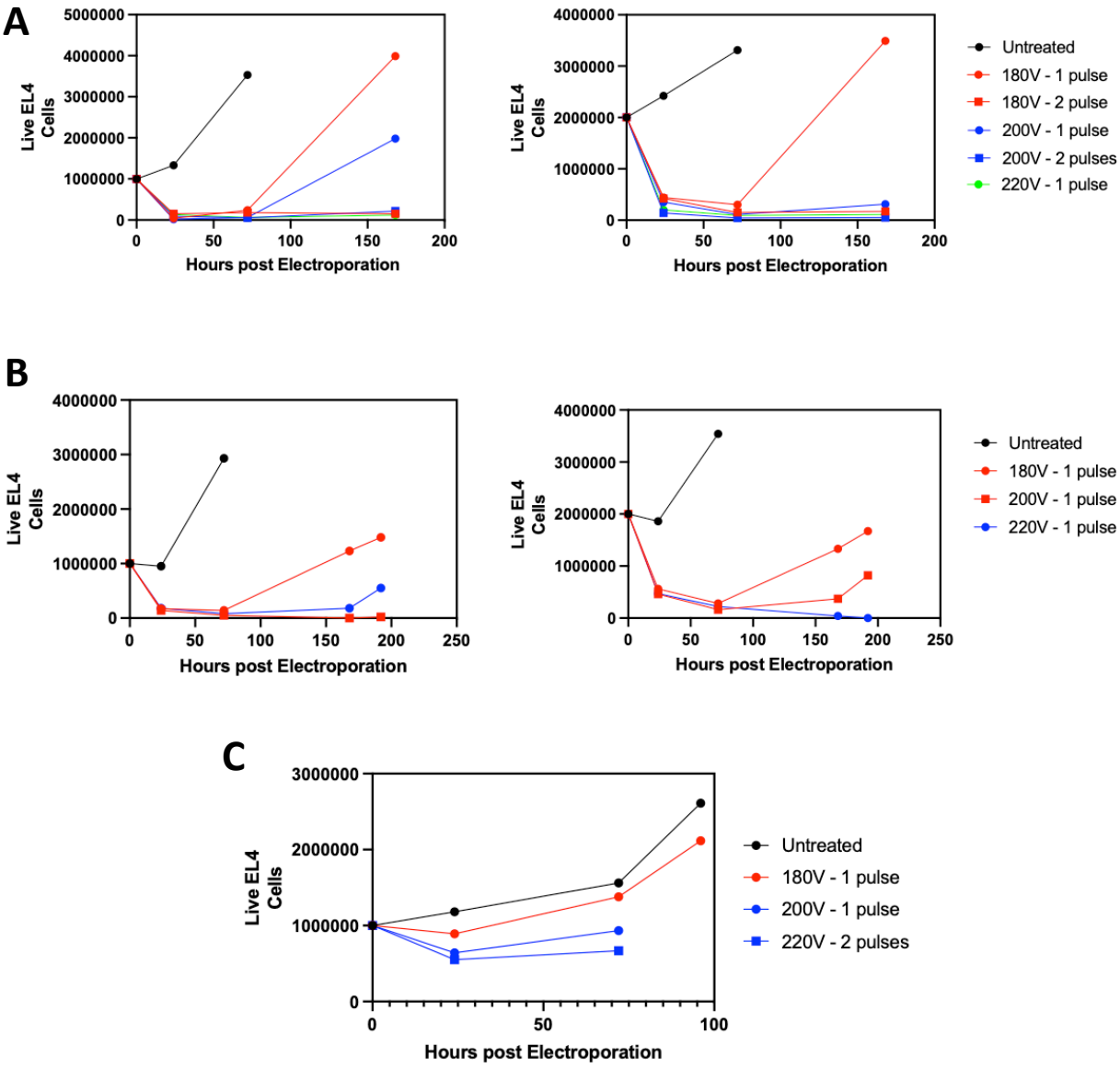


*Figure 17. Optimisation is required to sustain acceptable El4 viability shortly after electroporation.*

Viability of EL4 cells when electroporated without CRSIPR-Cas9 RNPs (A) using a 0.1cm (B) or 0.2cm Gene Pulser XCell microcuvettes at either 1x106 or 2x106 EL4 cells per cuvette under specified electroporation voltage conditions. (C) Viability of EL4 cells under improved practical conditions using a 0.2cm cuvette of which data points are the mean of several electroporation experiments of the same voltage conditions which included RNPs (crRNAs used – SRSF1-AA, SRSF1-AC, NTC).

within electroporated EL4 cell populations (Figure 18a). A subsequent transfection at 220V for 2 pulses and the same RNP conditions achieved a detectable loss of SRSF1 protein abundance (Figure 18b). The largest decrease of SRSF1 occurred when combining the two SRSF1 targeting crRNA associated RNPs. These results showed that transfection of *Srsf1* targeting crRNAs containing CRISPR-Cas9 RNPs could reduce the abundance of functional SRSF1 protein within EL4 cells.

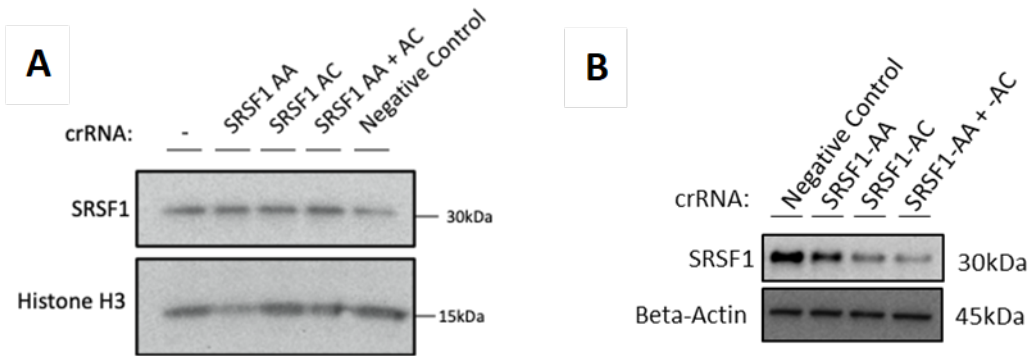


*Figure 18. Electroporation mediated transfection of SRSF1 targeting CRIPSR-Cas9 RNPs into EL4 cells reduces SRSF1 abundance*

Western blots showing resulting SRSF1 protein abundance at (A) 96 hours and (B) 72 hours post-electroporation of EL4s with RNPs *(*1µM*)* associated the following crRNAs: SRSF1-AA, SRSF1-AC and NTC, at (A) 180V, 1 pulse and (B) 220V, 2 pulses.

Since this method of transfection yielded less than 100% efficiency, electroporated EL4 populations were serially diluted and expanded in an attempt to gain clonal populations completely negative for SRSF1 expression. Following the first round of serial dilution, 24 populations were screened. From these no SRSF1 negative clonal populations were identified (Figure 19a). Two populations (AAAC1_G5 + AAAC2_E7) with low SRSF1 expression relative to negative control crRNA treated EL4 cells were selected for a further round of serial dilution and expansion. One population with relatively high SRSF1 expression (SRSF1AAAC1_E6) was also diluted and expanded to gain a positive control for further screens. 24 further populations were selected for screening; however, 6 of these did not expand once being taken out of the 96 well dilution plate during the course of experiments. Of these populations all still had detectable SRSF1 protein (Figure 19b, showing 10 population of 18 screened). Initial screens had been taken approximately 3-5 days following removal from 96 well plates without disturbance to passage or add media. Further screens of the same populations following subsequent 72hr passages (Figure 19b, Screen 2-3) showed that SRSF1 protein abundance returned to an increased and more stable level (AAAC_G5_B7, Screens 1-3). This suggests that the proportion of SRSF1 negative cells in a heterogeneous EL4 population is not maintained over time.

### 3.3.3 SRSF1 knockdown in EL4 cells

Alongside work using CRISPR-Cas9 to attempt to knockout SRSF1 within EL4 cells, we also tested an antisense-oligonucleotide gapmer to instead knockdown levels of SRSF1. Gymnosis

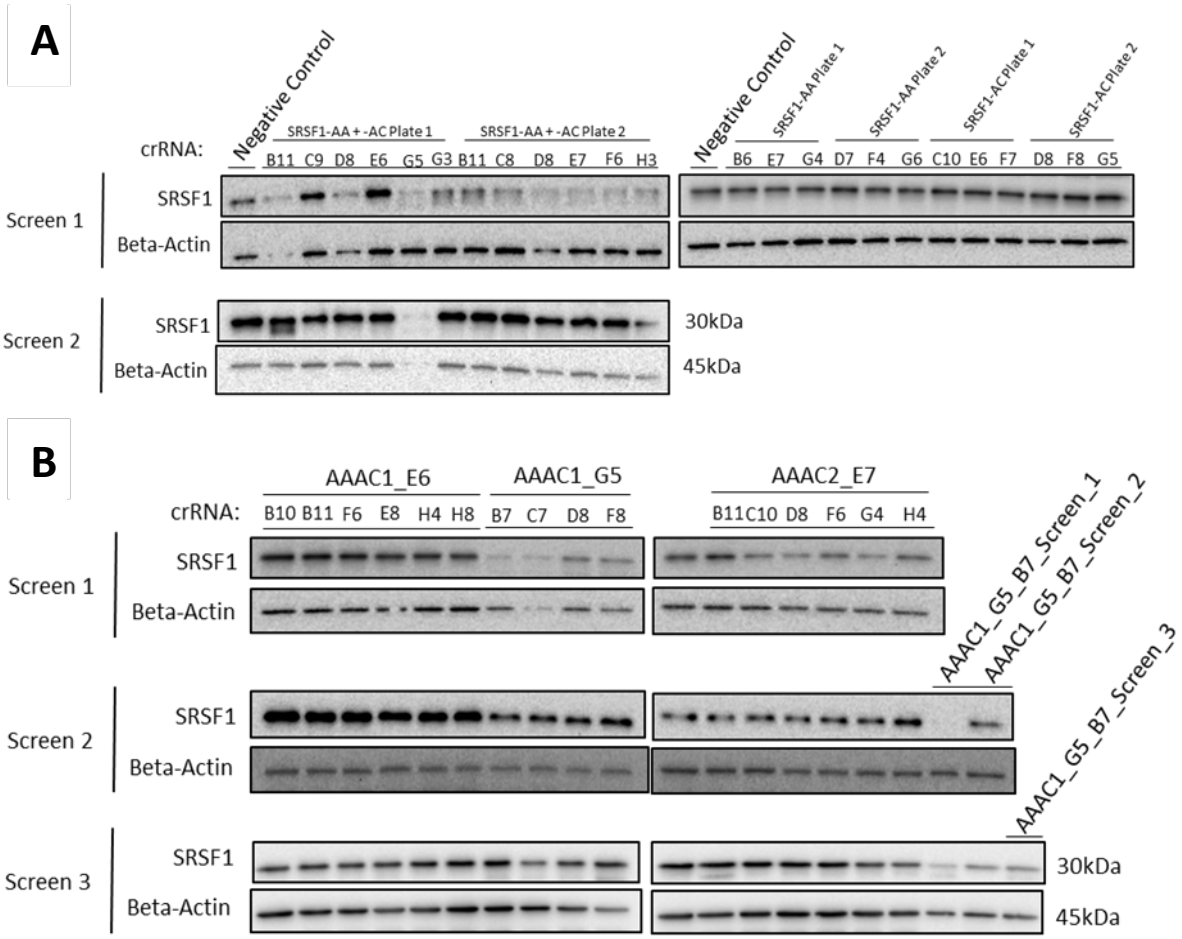


*Figure 19. Cloning by limiting dilution did not produce a clonal SRSF1 negative population.*

(A) SRSF1 protein abundance in resulting EL4 populations following serial dilution and expansion of electroporation transfected populations from Figure 18b. (B) SRSF1 protein abundance in resulting EL4 populations following serial dilution and expansion of populations from (A).

is the process by which gapmers enter cells unassisted (Stein, et al., 2010). Lipofection, using Lipofectamine™ 2000, and electroporation, at two differing conditions using the Gene Pulser Xcell system (220V-1 pulse, 220V-2 pulses), were also used to assist more efficient transfection of the SRSF1 gapmer into EL4 cells. Results showed that at 72 hours post-transfection there was a slight reduction in total SRSF1 protein in EL4 cells treated with the SRSF1 gapmer only where no transfection assistance was used (gymnosis) (Figure 20a). The proposed mechanism of the SRSF1 gapmer is binding to its complementary sequence within *Srsf1* mRNA for RNase H1 mediated degradation of LNA-DNA duplexes (Liang, et al., 2017). To confirm this, levels of SRSF1 mRNA was quantified using RT-qPCR upon *Srsf1* targeting gapmer treatment. Results showed that both gymnosis and lipofection-assisted transfection of the SRSF1 gapmer lead to reduction (~20-25%) of SRSF1 expression (Figure 20b). Electroporation

assisted gapmer transfection reduced neither SRSF1 protein nor RNA abundance. Overall, the SRSF1 gapmer was successful at a small reduction of SRSF1 protein within EL4 cells, however likely not to a standard where the effect of SRSF1 on post-transcriptional regulation could be investigated.

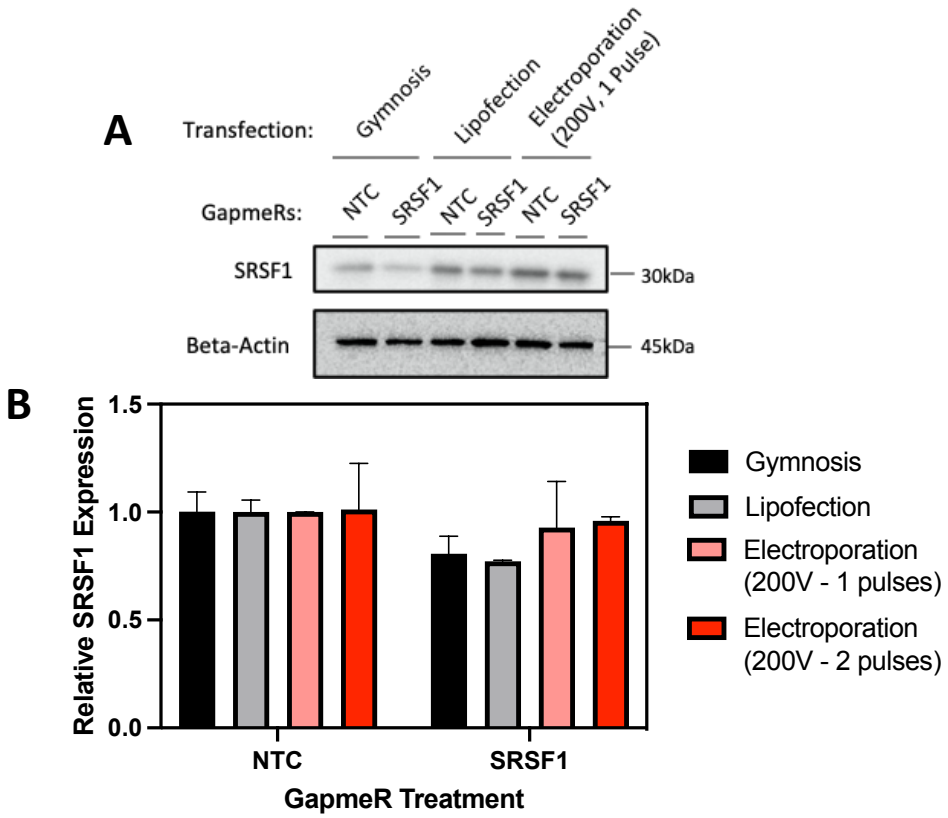


*Figure 20. Gymnotic transfection of SRSF1 targeting gapmer into EL4 cells causes a knockdown of SRSF1 protein.*

(A) Western blot showing the SRSF1 abundance in EL4 cells treated with either NTC and SRSF1 targeting gapmers taken up by unassisted gymtotic transfection, lipofection with Lipofectamine ™ 2000 transfection reagent or electroporation with the Gene Pulser Xcell. (B) qPCR results showing relative SRSF1 expression following the same *g*apmers treatment as aforementioned in EL4 cells. All expression values are relative to the respective NTC value in each transfection condition.

### 3.3.4 SRSF1 knockout in naïve CD4+ T cells

Having observed the effect of SRSF1-AA and SRSF1-AC crRNA on the transient depletion of SRSF1 in EL4 cells, we planned to use the same CRISPR-Cas9 mediated methodology in naïve CD4+ T cells. As we were particularly interested in investigating the effect of SRSF1 on transcription and alternative splicing across CD4+ T cell differentiation it is important that the CRISPR-Cas9 mediated SRSF1 knockout event occurs within naïve CD4+ T cells, prior to the activation stimuli.

Prior to *in vitro* activation stimuli, naïve CD4+ T cells were electroporated along with CRISPR-Cas9 RNPs using the Lonza 4D-Nucleofector® X Unit under previously optimised settings (Seki & Rutz, 2018). A non-electroporated condition (standard *in vitro* activation) was used alongside SRSF1 targeting/negative control crRNA RNP treated cells to ensure that electroporated of naïve CD4+ T cell did not negatively affect their ability to differentiate into Th2 cells following the appropriate signals. Following electroporation, CD4+ T cells saw a drastic reduction in viability with their populations reducing by ~75% 4 days post activation/electroporation compared to the initial cell input at day 0 (Figure 21a). Despite this, quantification of Th2 hallmark cytokines by flow cytometry at 6 days post activation and electroporation showed the percentage of IL10 and IL4 positive CD4+TCR+ cells were significantly higher than under the non-electroporated conditions (Figure 21b,c). This suggests that electroporation, whilst reducing cell viability, does not negatively affect differentiation of CD4+ T cells into Th2 cells *in vitro*. Additionally, whilst our results actually showed that electroporated CD4+ T cell populations had a significantly higher percentage of IL4 and IL10 positive cells, this is not conclusive evidence to show that electroporation improves activation. This is likely due to increased input cell numbers of electroporated SRSF1 targeting and NTC conditions compared to the non-electroporated control in an attempt to account for any cell death that occurred post-electroporation (Figure 21a).

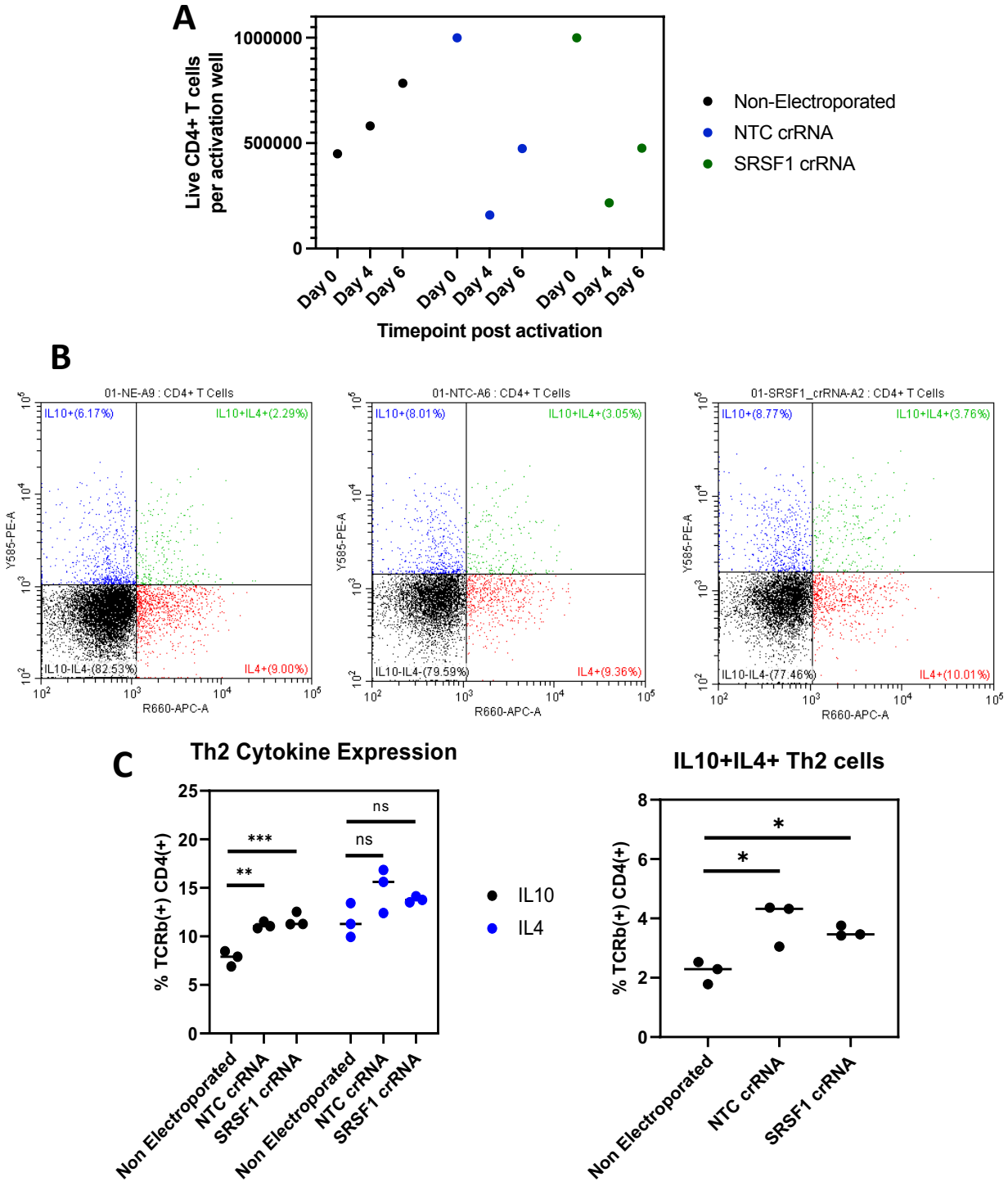


*Figure 21. Electroporation of CD4+ T cell does not negatively affect expression of IL4 and IL10 following Th2 polarisation in vitro.*

(A) Viability of CD4+ T cells following electroporation. (B) Model flow cytometry plots showing the percentage of IL10+ (Y585-PE, TL quadrant), IL4+ (R660-APC, BR quadrant) and IL10+IL4+ within CD4+ T cell populations 6-days post activation with anti-CD3. Conditions differed by non-electroporated, electroporated with NTC crRNA associated RNPs and electroporated with SRSF1 crRNA associated RNPs. (C) Plots showing the percentage of single positive IL10 and IL4 and double positive IL10, IL4 CD4+ T cells. One-way anova and subsequent multiple comparison test determined statistical significance of results. Concerning P values: *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$.

SRSF1 protein abundance was analysed at both 4- and 6-days post activation/electroporation. In naïve CD4+ cell populations electroporated alongside SRSF1-AA/AC associated RNPs there was no detectable change in SRSF1 protein abundance at either timepoint compared to the non-electroporated Th2 cells (Figure 22). Therefore, under the conditions used, this method of RNP transfection into primary CD4+ T cells was unsuccessful.

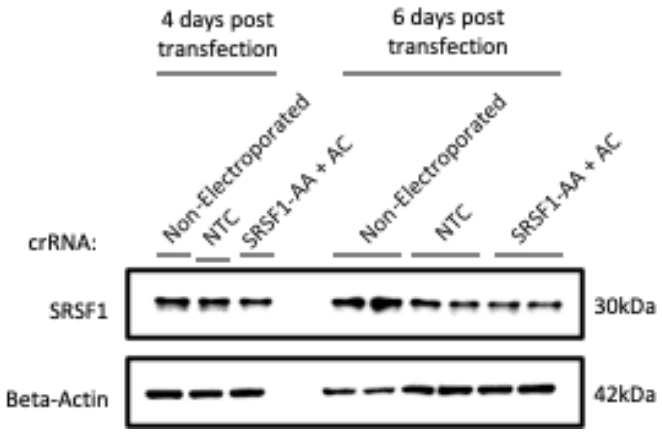


*Figure 22. Transfection of SRSF1 targeting CRISPR-Cas9 RNPs to reduce SRSF1 abundance in Th2 cells was unsuccessful.*

Western blot showing SRSF1 abundance 4 days (prior to IL2 treatment) and 6 days post electroporation with CRISPR-Cas9 RNPs (2µM) and *in vitro* activation with anti-CD3.

## 4. Discussion

This project aimed to investigate the regulatory output of *Malat1* on SRSF1 function and subsequently better characterise the molecular mechanisms of this regulatory action during Th2 polarised CD4+ T cell differentiation. Therefore, a bioinformatic analysis and comparison of iCLIP data was carried out to identify a group of transcripts differentially bound by SRSF1 in *Malat1*$^{-/-}$ Th2 cells compared to *WT*. *In vitro* validation of regulation of two selected transcripts by *Malat1* was carried out within Th2 cells. This revealed that, in addition to increased SRSF1 enrichment of *Runx3*, the loss of *Malat1* leads to both a reduced general abundance of *Runx3* and lower isoform usage of *dRunx3* in late stage differentiation Th2 cells. Further *in vitro* validation is required for several other identified CD4+ T cell relevant candidate transcripts.

Already having access to validated and efficient *Malat1*$^{-/-}$ *in vitro* models within the Lagos lab, we also aimed to produce a CRISPR-Cas9 mediated *Srsf1*$^{-/-}$ *in vitro* model for investigation of the splicing factor's regulatory action across CD4+ T cell differentiation. SRSF1 targeting crRNA

associated RNPs were successfully shown to transiently reduce SRSF1 protein abundance within EL4 cells. However, a complete and stable knockout of SRSF1 within both the EL4 cell line and primary naïve CD4+ cells was unsuccessful and requires further optimisation.

## 4.1 *Malat1* regulates SRSF1 binding to regulators of Th2 cell differentiation

CLIP is an established technique used for defining the RNA interactome of a desired protein via quantification of crosslink sites. There are a large number of these techniques utilising crosslinking, which include photoactivatable ribonucleoside-enhanced CLIP (PAR-CLIP) and CrossLinking, Ligation, and Sequencing of Hybrids (CLASH). These techniques possess different uses as well as limitations; for example, PAR-CLIP is able to offer improved crosslink efficiency and reduced background noise due to its introduction of photoactivatable nucleoside 4-thiouridine into cells prior to crosslinking (Hafner, et al., 2010). CLASH is able offer insight into RNA-RNA interactions mediated by RBPs (Helwak, et al., 2013). CLIP methods have been vital for development of research characterising the important role of RNPs in developmental, neurological and immunological contexts (Hafner, et al., 2021) (Díaz-Muñoz & Turner, 2018). Other advanced approaches to assess RNA-RBP interactions include targets of RNA-binding protein identified by editing (RBP-Tribe) and STAMP (Surveying Targets by APOBEC-Mediated Profiling), which offer advantages over crosslinking methods through a dynamic temporal view of an RBPs RNA interactions using very low cell numbers (RBP-Tribe) and improved view of RBPs bound to a single RNA molecule (STAMP) (McMahon, et al., 2016) (Brannan, et al., 2021).

Our approach uses iCLIP, which attains precise resolution of crosslink sites due to an intramolecular cDNA circularization step prior to sequencing. This allows for specific identification of the truncation points that were present during reverse transcription of RNA transcripts, giving single nucleotide resolution of protein-RNA crosslink sites which are mapped across transcriptomes (Hupertz, et al., 2014). Our bioinformatic analysis of iCLIP data to assess changing SRSF1 crosslink signal of transcripts, namely those with a role in CD4+ T cells, across *Malat1* expression conditions in Th2 cells aimed to identify candidate transcripts through which *Malat1* mediated SRSF1 regulation of CD4+ T cell differentiation occurs.

Whilst this study does only consider the use of one technique, iCLIP, the wealth of other techniques briefly mentioned above to assess and unravel the dynamics of RBP-RNA

interactions within the cell could be deployed in future experimentation. Specifically, TRIBE-STAMP could be used alongside iCLIP to better characterise and understand the dynamics of SRSF1-RNA interaction within Th2 cells across different activation timepoints; additionally giving a view of other RBPs bound to RNA targets identified to allow for a more mechanistic view of the action of SRSF1 binding (Flamand, et al., 2022).

Through our analysis of the general structure of the Th2 SRSF1-iCLIP dataset we could ascertain its general validity compared to other published SRSF1 CLIP studies. 98% of transcripts found within the SRSF1 CLIPdb were also displayed SRSF1 crosslink signal within *Malat1*$^{-/-}$ and WT Th2 cells (Yang, et al., 2015). Several of our detected SRSF1 bound targets were also seen within late stage thymocytes. For example, *Runx3*, *Runx1*, *Icos*, *Stat1* and *Il27a* are common CD4+ T cell relevant SRSF1 bound transcripts between Th2 cell and late stage thymocytes (Qi, et al., 2021). Common RNA targets of SRSF1 between varying contexts increases confidence in our findings. The distribution of SRSF1 crosslink events in pre-mRNA exonic and intronic sequences found within our study, however, was not consistent with late stage thymocytes. Within our Th2 cells, SRSF1 exhibits a preference for binding exonic sequences, while within late-stage thymocytes, SRSF1 shows a preference for intronic sequences (Qi, et al., 2021). SRSF1 is able to carry out splicing regulation by binding to both exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs), while also participating in competitive binding with hnRNPs at exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs) (Motta-Mena, et al., 2010) (Lynch & Motta-Mena, 2013) (Jobbins, et al., 2021). Therefore, whilst this variability in exonic/intronic binding preference displays that SRSF1 binding behaviour can be context-dependent, with its role in splicing regulation being influenced by different cellular contexts, it does not imply a weaker function as a splicing regulator. Instead the variability highlights further the multifaceted roles of SRSF1 in splicing.

Several CLIP based studies of SRSF1 across different contexts, including human embryonic kidney cells, MCF7 breast cancer cells and late stage thymocytes have shown a purine rich consensus sequence for SRSF1 binding (Sanford, et al., 2009) (Du, et al., 2021) (Qi, et al., 2021). A limitation of our own analysis was that it did not include a genome wide analysis of SRSF1 binding site nucleotide preferences. One way we could have completed this would be to use software, such as MEME suite (Bailey, et al., 2015), to analyse the SRSF1 iCLIP crosslink data to generate the proportion of each nucleotide making up all single nucleotide SRSF1-

RNA binding crosslink sites. From this, we would expect to see the purine nucleotides (guanine and adenine) making up a greater proportion of crosslink sites, to align with previously published SRSF1 consensus sequences.

Our analysis revealed a positive correlation between the abundance of a transcript and its SRSF1 crosslink signal, suggesting that transcripts with higher expression levels tend to exhibit elevated SRSF1 binding. This observation could imply that SRSF1 association increases due to the increased availability of binding sites as transcript abundance is higher. This interpretation agrees with published evidence of *Malat1* and SRSF1 as binding partners. Since its discovery *Malat1* has been noted as having ubiquitous and relatively high abundance compared to other ncRNAs and even most protein coding mRNAs (Ji, et al., 2003). For example, our RNAseq data of *WT* Th2 cells shows *Malat1* top highest expressed transcript in the cell; with a counts per million reads mapped (CPM) value of 16,873 (1.68% of cellular transcripts). Within WT Th2 cells, *Malat1* also has the highest SRSF1 crosslink signal of any single transcript (30,399 Crosslink Events per million). This could be a result of its relatively high abundance. However, it is also crucial to acknowledge that SRSF1 binding of transcripts contributes to changes in transcript abundance. Therefore, the positive correlation seen between abundance and SRSF1 crosslink signal should not be interpreted as a complete regulatory redundancy of RNA binding by SRSF1. This point is supported by our results indicating the drastic change of SRSF1 binding on several important CD4+ T cell regulators upon *Malat1* loss. This shows that despite the correlation between *Malat1* abundance and SRSF1 binding there is indeed important *Malat1* mediated riboregulation of SRSF1 RNA binding behaviour occurring in Th2 cells.

Due to the influence of abundance on SRSF1 crosslink signal of transcripts normalisation of SRSF1 crosslink signal was carried out. This was important as it allowed identification of transcripts which, whilst their expression changes significantly, their $\log_2$(csFC) does not. For example, *Bcl2*, an important apoptotic regulator implemented in CD4+ T cell activation (Yang, et al. 2017), has a $\log_2$(csFC) of 0.15. Yet the expression of Bcl2 has a 2-fold decrease in *Malat1*$^{-/-}$ Th2 cells compared to WT conditions. Therefore, the normalised $\log_2$(csFC) value of *Bcl2* sits much higher at 1.08. Whilst not initially appearing so, following normalisation of $\log_2$(csFC) values to expression this data shows that *Malat1* regulates SRSF1 binding to *Bcl2*.

Whilst not especially considered during this study, false-positive RNA binding targets of SRSF1 via non-specific binding during iCLIP could compromise the reliability of data. To ensure this

reliability in RNA-protein interaction data generated from iCLIP, control measures include use of an isotype negative control antibody as well as a no-antibody control during the immunoprecipitation step of iCLIP. These steps remove background noise or false-positive interactions caused by non-specific binding of cellular protein to the anti-SRSF1 antibody. More stringent methods of control include RNase treatment following immunoprecipitation of SRSF1 to remove any non-crosslinked RNA from the sequencing step and a non-crosslinked control lacking UV treatment to isolate binding arising specifically from crosslinking (Hupertz, et al., 2014). As iCLIP and other surrounding methods develop, methods of controlling the specificity of RNA-RBP interactions improve, presenting an opportunity for our future experiments to enhance confidence in results by incorporating improved control steps. This said, it is certainly a limitation of the study that there is a lack of enhanced considered for false-positive SRSF1-RNA interactions.

To summarise, over the course of this whole bioinformatic analysis, we identified a large group of candidate transcripts which showed a greater than 1.5-fold increase or decrease in SRSF1 crosslink signal in *Malat1*$^{-/-}$ Th2 cells (Appendix 1.). Whilst the bioinformatic analysis identified a large number of transcripts alternatively bound by SRSF1 in Th2 cells upon *Malat1* loss, all of which deserve specified investigation and validation looking ahead to future work, only two could be investigated further within this project, *Il2ra* and *Runx3*.

## 4.2 *Malat1* does not regulate *Il2ra* abundance through altered usage of the alternatively spliced *Il2ra* truncated isoform

IL2RA is an extracellular component of the IL2 receptor (Minami, et al., 1993) which has the function of binding the IL2 cytokine, an essential event for correct expression of effector cytokines and cell proliferation in CD4+ T cells (Ross & Cantrell, 2018). Mutations associated with both loss and overexpression of *Il2ra* within CD4+ T cells are associated with autoimmune diseases such as follicular bronchiolitis, IPEX and colitis (Bezrodnik, et al., 2013) (Caudy, et al., 2007) (Joosse, et al., 2021), stressing the importance of correct *Il2ra* regulation in CD4+ T cell activation. As shown previously in the Lagos lab, *Malat1* positively regulates IL10 expression in CD4+ T cells (Hewitson, et al., 2020). Interestingly, the IL2 response has been linked to increased IL10 expression in Tregs (Zhou, et al., 2021) and IPEX patients with low expression of *Il2ra* additionally have low IL10 expression in their CD4+ T cells (Caudy, et al., 2007). Therefore, both *Malat1* and *Il2ra* are vital regulators of both CD4+ T cells activation

and IL10 expression as well as having their dysregulation linked with several autoimmune disorders (Yang, et al., 2017). As such, our investigation aimed to investigate the regulatory function of *Malat1* on *Il2ra* abundance and isoform usage within Th2 cells, following evidence of *Malat1* regulated SRSF1 binding to *Il2ra* at the truncated isoform splice site.

We hypothesised that *Malat1* regulates *Il2ra* expression across Th2 differentiation through indirect regulation of SRSF1 mediated alternative splicing of *Il2ra* into the truncation isoform. This proposed mechanism relies on truncated *Il2ra* producing a non-functional protein, leading to aberrant CD4+ T cell activation and hallmark cytokine expression due to loss of IL2 signalling (Minami, et al., 1993). Whilst no published evidence exists showing this, truncated *Il2ra* only consists of exon 1, showing a loss of 93% of canonical pre-mRNA sequence. Additionally, the amino acids reported to form direct molecular interactions with IL2 are coded by codons within exons 2-8 of *Il2ra* mRNA (Stauber, et al., 2006). Therefore, we were confident to predict that the protein produced by truncated *Il2ra* was non-function as a component of the IL2 receptor or even in a soluble form (Maier, et al., 2009) due to the lack of possible IL2 binding.

There is no published evidence showing that truncated *Il2ra* exists transcript within Th2 cells, therefore we showed using primers specific for truncated *Il2ra* that this transcript is indeed present. Our results suggested that *Malat1* does not regulate total *Il2ra* abundance through SRSF1 mediated regulation of the alternative splicing event producing truncated *Il2ra* in Th2 cells. Despite this, the effect of *Malat1* loss on the SRSF1 binding peak at the truncated *Il2ra* splicing site was extremely prominent and the potential of its possible effects should not be understated. To be truly conclusive that *Malat1* mediated SRSF1 binding to *Il2ra* is not altering *Il2ra* expression in any way, further experimentation is required.

For example, regulation of splicing and stability of mRNA transcripts is often carried out by several different RBPs and can be competitive (Motta-Mena, et al., 2010). This is also true for *Il2ra*. One study shows that the RBP and splicing factor HuR exerts a pivotal role in the post-transcriptional regulation of *Il2ra* by binding to its 3' UTR and enhancing mRNA stability. A loss of HuR leads to suboptimal *Il2ra* translation and a weaker Th2 activation and differentiation (Techasintana, et al., 2017). This shows that *Il2ra* is post-transcriptionally regulated through an RBPs and along with our result, showing SRSF1 binding it is likely *Il2ra* is bound by several RBPs. To gain a comprehensive summary of the interactome of *Il2ra*, RAP-

MS could be carried out within Th2 cells and would provide better insight into the post-transcriptional regulation of *Il2ra*. This approach could provide an explanation for the lack of effect on abundance or isoform usage following the decrease of SRSF1 binding upon *Malat1* loss. For example, SR proteins in some cases have been shown to be redundant within their binding sites and effect on bound transcript RNA metabolism (Liu, et al., 1998) (Howard & Sanford, 2016). Therefore, to consider the whole interactome of a candidate transcript is a vital approach when concluding the regulatory effect of just one RBP. Which in this case, is the effect of *Malat1* regulated SRSF1 binding of *Il2ra*.

Additionally, there is increasing evidence of the importance of the cytoplasmic function of SRSF1 (Tripathi, et al., 2010) (Haward, et al., 2021) and evidence that SRSF1 additionally can regulate the translation of its mRNA binding partners (Maslon, et al., 2014). As well as this, *Malat1* has been reported to carry out indirect regulation of translation through characterised translation regulators certain cancer-related contexts (Malakar, et al., 2019) (Ji, et al., 2019). This is a mode of action that *Malat1* could be exhibiting upon interaction with SRSF1, altering translation of SRSF1 target transcripts. Without considering the wider effects of *Malat1* and SRSF1 outside of post-transcriptional control, no validation of candidate transcripts regulation by *Malat1* and/or SRSF1 should be considered complete without both assessing mRNA and protein abundance. Therefore, analysis by Western Blot or Flow Cytometry of IL2RA levels on Th2 surface following *Malat1* loss would be an essential follow up experiment to this project.

### 4.3 *Malat1* regulates *Runx3* expression in late stages of Th2 activation

*Runx3* is a major transcription factor and regulator of both Th1 and Th2 cell differentiation and a previously reported RNA binding partner of SRSF1 (Qi, et al., 2021). It is also characterised as a positive regulator of CD8+ T cell activation (Wong, et al., 2011). In the context of Th cell differentiation, it inhibits GATA3 activity whilst promoting T-BET activity ( Korinfskaya, et al., 2021). *Runx3* therefore, is expressed at higher levels in Th1 than Th2 cells. Whilst Runx3 inhibits GATA3, some *Runx3* expression is hallmark feature for Th2 cell activation, as a complete lack of *Runx3* expression is characteristic of the naïve CD4+ T cell phenotype (Egawa, et al., 2007) (Naoe, et al., 2007). However, induced overexpression of *Runx3* in naïve CD4+ T cells activates IFNy, whilst silencing IL4 expression in polarised Th2 cells. Further aberrant *Runx3* overexpression causes *in vitro* polarisation assays to only yield

Th1-like IFNy expressing cells, despite canonical Th2 polarising conditions (rIL-4 treatment) (Kohu, et al., 2009). Furthermore, mice with T cell specific deficiency of *Runx3* are prone to asthma, a disease associated with aberrant Th2 response (Naoe, et al., 2007). This evidence emphasises the importance of dynamic regulation of *Runx3* abundance between Th1 and Th2 cells to maintain the required phenotype.

The *Runx3* gene can be transcribed from the distal or proximal promotor, producing the *dRunx3* and *pRunx3* transcripts respectively. *pRunx3* contains an inefficient kozak sequence within exon 1, leading to poorer translation of *pRunx3* compared to *dRunx3* (Kim, et al., 2015). This is used as a transcriptional mechanism of control within several contexts. Promotor specific *Runx3* expression regulates linage selection of thymocytes (Egawa, et al., 2007). Following positive selection in the thymus, *dRunx3* expression alone specifically defines the presence of increased RUNX3 protein in CD8+ T cells (Littman & Egawa, 2008) and the absence of RUNX3 within naïve CD4+ T cells is defined by increased *pRunx3* expression and repression of *dRunx3* (Egawa, et al., 2007). This alternative promotor preference of *Runx3* also defines effector Th cell phenotype and cytokine expression. *dRunx3* is expressed at significantly higher levels in Th1 cells compared to Th2 cells, whilst *pRunx3* is of similar levels (Kim, et al., 2015). This leads to higher *RUNX3* protein in Th1 compared to Th2 cells at 6-days post activation (Kohu, et al., 2009).

We showed that SRSF1 and *Runx3* binding, a previously detected RNP interaction in thymocytes (Qi, et al., 2021), increases upon *Malat1* loss in Th2. Additionally, SRSF1 displays a specific binding preference for *dRunx3* which is apparent at the 3' and 5' UTR, an effect enhanced by *Malat1* loss. SRSF1 has the function of regulating mRNA stability through binding with UTRs, increasing or decreasing stability upon binding in a transcript dependent manner (Paz, et al., 2021). Therefore, we were keen to investigate whether this *Malat1* regulated SRSF1 binding effects isoform usage of distal/proximal *Runx3* as a molecular mechanism to regulate Th2 differentiation.

Immediately following activation from naïve CD4+ T cells to Th2 cells, *Malat1* has a characterised downregulation (Hewitson, et al., 2020) whilst *Runx3* is upregulated (Kim, et al., 2015). Over the course of differentiation, *Runx3* expression is strictly maintained as overexpression can lead to a loss of Th2 phenotype of cytokine expression. Our results characterise *Malat1* as an up regulator of *Runx3* through directed upregulation of the

translationally efficient *dRunx3*, at later timepoints indicative of differentiation (5-6 days post activation). Therefore, the downregulation of *Malat1* upon activation of Th2 cells (Hewitson, et al., 2020) is likely vital for preventing aberrant overexpression of *Runx3* during Th2 differentiation. At earlier timepoints, indicative of activation (0-24 hours post activation), however, *Malat1* appears to have no control over *Runx3* abundance, only upregulating the translationally inefficient *pRunx3*. This timepoint specific effect of *Malat1* on *Runx3* is likely important as the *Malat1* loss and *Runx3* upregulation shortly following activation would be contradictory to findings of *Malat1* positive regulation of *Runx3* at later timepoints. Western Blot or flow cytometry experiments concurrent to this analysis would have been beneficial to evaluate RUNX3 protein abundance across these timepoints.

*Malat1* loss immediately following activation is also an important characteristic of Th1 activation, of which increased *Runx3* expression compared to that of Th2 cells is vital for positive regulation of Th1-phenotype driver T-BET. This again, is contradictory of our findings of *Malat1* positive regulation of *Runx3* during differentiation. This, along with the timepoint specific effect of *Malat1* on *Runx3* isoform usage is indicative of a much more complex regulatory mechanism surrounding these two regulators across the whole CD4+ T cell differentiation trajectory. Further investigation, including extensive repeats of RT-qPCR analysis of *Malat1* loss on *Runx3* abundance and isoform usage, is essential for validation the above speculation. Positive validation could confirm a previously unpublished regulatory axis between *Malat1* and *Runx3*, which due to the balance of *Runx3* expression regulating Th1 and Th2 phenotype, additionally would assign *Malat1* as a regulator of Th phenotype plasticity.

As previously stated, *Malat1* upregulates IL10 expression in Th1 and Th2 cells (Hewitson, et al., 2020). No direct link currently exists characterising *Runx3* as a direct IL10 regulator during CD4+ T cell activation. However interestingly, in type 2 Innate Lymphoid Cell (ILC2s), which act with a Th2-like phenotype and cytokine expression profile, *Runx3* depletion causes an increase in IL10 expression (Miyamoto, et al., 2019). Therefore, there lies some plausibility within *Runx3* acting as a molecular mediator for characterised SRSF1 and *Malat1* regulation of correct CD4+ T cell activation and Th phenotype.

Further experimentation would also be beneficial for investigation of the role of SRSF1 in this proposed regulatory axis. Linking increased SRSF1 binding to *Runx3*, specifically at the UTRs, to an effect on mRNA stability upon *Malat1* loss would require naïve CD4+ T cell *Srsf1*$^{-/-}$ for *in*

*vitro* Th2 differentiation. Within this model we would expect a loss of SRSF1 to cause upregulation of *Runx3* regardless of *Malat1* expression. This would be resultant of no SRSF1 binding to *Runx3* UTRs, causing an increase to *Runx3* mRNA stability, specifically in *dRunx3* at later Th2 differentiation timepoints. To specifically investigate an effect on mRNA stability, a transcriptional inhibitor such as Actinomycin D could be used to halt transcription (Ratnadiwakara & Änkö, 2018). Subsequent RT-qPCR analysis of the rate of total *Runx3* and promotor specific isoform abundance decrease across defined timepoints following transcriptional halting between two SRSF1 expression conditions in Th2 cells would reveal whether SRSF1 was responsible for regulating *Runx3* stability.

## 4.4 Electroporation mediated transfection of SRSF1 targeting crRNA associated RNPs reduces SRSF1 abundance in EL4 cells.

*In vitro* models can be used to validate transcriptomic or proteomic regulation by a molecular target through use of established knockout/knockdown methods. The Lagos lab has established methods for knocking down *Malat1 in vitro* using *Malat1* targeting oligonucleotide gapmers. These work with an efficiency of 80% within EL4 cells after 72 hours (data not shown) and up to 90% in primary CD4+ T cells (Hewitson, et al., 2020). As well as this, naïve CD4+ T cells have been taken from *Malat1*$^{-/-}$ mice (Nakagawa, et al., 2012) and used within *in vitro* activation assays to investigate the effect of *Malat1* on CD4+ T cell activation and differentiation. To complement these *Malat1* focussed models within this project we were keen to develop CD4+ T cell relevant *in vitro* knockdown/knockout models of SRSF1.

CRISPR-Cas9 is an increasingly more accessible method of targeted gene editing that can be used for addition of genetic material as well as gene knock-ins or knockouts (Ran, et al., 2013). We used transfection of CRISPR-Cas9 RNPs into cells to induce directed double strand cleavage. RNPs were selected due to evidence of their improved knockout efficiency and viability in T cells, compared to a plasmid orientated approach (Oh, et al., 2022; Su, et al., 2016). CRSIPR-Cas9 RNP complexes are formed of the Cas9 nuclease associated with a small guide RNA (sgRNA) consisting of tracrRNA linking it to the Cas9 nuclease and a CRISPR-RNA (crRNA) which be specifically designed to target to a desired sequence in the genome. This directs the Cas9 enzyme to cleave both the sense and antisense DNA strand and prompts non-homologous-end-joining (NHEJ), a natural cellular process, to repair this cleavage. NHEJ is an

error-prone process, which is advantageous for gene knockouts as indels occurring within a coding exon can lead to frameshift mutations and premature stop codons (Ran , et al., 2013).

There are several methods of CRISPR-Cas9 RNP transfection methods, including electroporation, viral delivery and lipofection (Fajrial, et al., 2020). Electroporation was chosen to transfect CRISPR-Cas9 RNPs into target cells as it has emerged as a more efficient and less toxic transfection method compared to that of viral transfection and lipofection, especially in primary T cells (Seki & Rutz, 2018). Transfection via electroporation of CRISPR-Cas9 RNP into cell lines, when optimised, can be utilised to achieve >90% editing efficacy within the electroporated population (Gratacap, et al., 2020) (Wong, et al., 2019). Evidence shows that each cell type requires specific electroporation conditions for optimal transfection efficacy (Jordan, et al., 2008). Concerning EL4 cells, studies show that EL4 cells can successfully be transfected with CRISPR-Cas9 plasmids as well as RNPs via electroporation for both CRISPR-Cas9 gene knockout and CRISPR activation of genes (CRISPRa) respectively (Yan, et al., 2012). (Jensen, et al., 2021). However, unlike in Jurkat cells (Liang, et al., 2015), few published studies detail the optimised electroporation conditions for high viability and transfection efficiency EL4 cells specifically using the Gene Pulser Xcell system. Our efforts at following optimisation lead an EL4 cell viability of 70-90% compared to non-electroporated after 24 hours.

SRSF1-AA and SRSF1-AC, which we validated as effective crRNAs for CRISPR-Cas9 mediated depletion of SRSF1, were chosen as they targeted separate sequences within exon 1 of the *Srsf1* gene. This is an appropriate locus to target as any indels introduced here due to NHEJ following Cas9-medited cleavage will ensure that the introduced frameshift mutation will affect the majority of the upstream transcription. Additionally, these crRNAs target the Cas9 nuclease to cleave with the sequence coding for the RNA recognition motif 1 (RRM1) in the SRSF1 protein (Cléry, et al., 2021). The RRM1 is vital for conferring the RNA binding specificity and therefore the post-transcriptional regulatory role of SRSF1. Our results showed that transfection with RNPs associated with SRSF1-AA and SRSF1-AC crRNA together resulted in a greater loss of SRSF1 in EL4 cells. This agrees with several studies showing that transfection of multiple crRNA associated RNPs targeting the same gene significantly increases transfection efficiency (Zetsche, et al., 2017) (Ju, et al., 2019). For this reason, going forwards RNPs containing SRSF1-AA and SRSF1-AC were used in combination to increase subsequent SRSF1 knockout attempts.

Whilst high transfection and knockout efficiencies have been reported, RNP transfection cannot induce a 100% effective SRSF1 knockout within a population of EL4s. so cloning by limiting dilution was required. Observations following limiting dilutions suggested that a loss of SRSF1 was detrimental for viability of EL4 cells. *Srsf1*-cKO mice had improved proliferation and activation phenotypes when SRSF1 was lost in CD4+ T cells (Katsuyama, et al., 2019). This evidence suggests SRSF1 could be an inhibitor of EL4 T cell phenotype *in vitro*. However, it is also vital to consider the cancer phenotype of EL4 cells alongside their T cell phenotype. SRSF1 has been characterised as an oncogene which promotes aberrant tumour growth in several types of cancer (Du, et al., 2021) Additionally, induced SRSF1 depletion within the DT40 cell line (avian leukosis virus induced bursal lymphoma cells) has been shown to reduce genome stability and subsequently viability (Li & Manley, 2005). Therefore, it is plausible that SRSF1 is essential for immortalised cancer cell line viability, including that of EL4s. As SRSF1 knockout was transient within EL4 screened populations following limiting dilution, it is likely that this method of limiting dilution was not producing populations originating from single clones. Therefore, to better investigate the importance of SRSF1 in EL4 growth, this project would benefit from single cell sorting to increase confidence in the homozygosity of populations.

Due to these observations, partial SRSF1 knockdown within EL4 cells may be a better route to investigate the post-transcriptional regulation of the splicing factor within cells lines modelling an *in vitro* CD4+ T cell context. Results from this study showed that an antisense oligonucleotide LNA gapmer targeting SRSF1 within EL4s produced a suboptimal knockdown of SRSF1 protein and a ~25% reduction of SRSF1 expression when transfected through gymnosis. Other methods of SRSF1 knockdown, such as lentiviral transfected shRNA as well as siRNA both targeting *Srsf1* RNA transcripts, have achieved up to an 80% knockdown efficiency (Arif, et al., 2023) (Du, et al., 2021). Considering this, this SRSF1 targeting gapmer could be judged as unsuccessful at their given concentration and experimentation with other knockdown methods in EL4 cells should be considered.

## 4.5 Electroporation of naïve CD4+ T cells with CRISPR-Cas9 RNPs does not affect subsequent Th2 polarisation

T cells are characterised as hard-to-transfect cells (Rahimmanesh, et al., 2020). Viral transfection of plasmids into T cells has shown to be particularly ineffective, one study only gaining a maximum of 40% knockout efficiency in cultured mouse CD4+ T cells transfected

with a crRNA and Cas9 expressing vector construct (Seki & Rutz, 2018) (Cheng, et al., 2022). This said, electroporation has also garnered poor knockout viability as a transfection method for CRISPR-Cas9 plasmid constructs and RNPs in T cells (Gresch, et al., 2004) (Xu, et al., 2018). Several studies have shown, however, that electroporation can lead to a high knockout efficiencies and viability when T cells are transfected with CRISPR-Cas9 RNPs following optimisation. Optimal transfection efficacy for CRISPR-Cas9 RNP mediated knockout in T cells varies on the gene target and importantly, the activation state of the cells. Seki and Rutz observed that electroporation mediated RNP transfection of naïve CD4+ T cells resulted in little detected knockout (~5%) after 48hrs, along with extremely low viability. The same experiment was repeated, with *in vitro* stimulation (anti-CD3 and anti-CD28) occurring immediately after electroporation. This showed that the effect on viability had largely been negated and knockout efficiency for the same target was much higher (~60%). This showed that *in vitro* stimulation of TcR significantly improved RNP transfection and knockout efficiency following electroporation in T cells (Seki & Rutz, 2018). Due to this, within our attempts to produce SRSF1 negative Th2 cells, we opted to introduce activation stimuli immediately after RNP transfection via electroporation in naïve CD4+ T cells. This is advantageous as successful knockout allows for analysis of the function of SRSF1 across the CD4+ T cell differentiation trajectory. However, this method also has its limitations as residual protein may still be present during the early stage activation. To account for this exact limitation we faced, Seki and Rutz incubated RNP transfected naïve CD4+ in rIL7 for 5 days, after which reduced post-electroporation viability reduction was negated and increased knockout efficiency, which ranged from 40-80% with varying gene target (Seki & Rutz, 2018). This is due to IL7 action as a vital growth factor which can improve naïve CD4+ T cell survival and proliferation *in vivo* (Tan, et al., 2001). Whilst not applying this innovation to the project, it is definitely a consideration for use for optimisation of this method of SRSF1 knockout in naïve CD4+ T cells.

Without growth cytokine treatment, naïve CD4+ cells lack proliferative capacity *in vitro* and electroporation, however optimised, will inevitably have some effect on viability. Additionally, within the *in vitro* activation assay, stimulated cells receive the IL2 (proliferative) signal at 4 days post-activation. Therefore, immediate recovery after electroporation of activated CD4+ cells is not possible (Hedfors & Brinchmann, 2003). Even under normal non-

electroporation conditions cell death can occur within these initial 4 days. Despite this, we found that electroporation of naïve CD4+ T cells did not negatively affect subsequent Th2 polarisation. Our results agree with evidence showing that successful upregulation of *Foxp3* occurred from electroporated CD4+ T cells following stimulation with anti-CD3 (along with IL2 and TGF-beta). These results are indicative of successful and intended Treg activation and differentiation following electroporation of the naïve population (Seki & Rutz, 2018).

Unfortunately, western blot results showed that our CD4+ T cells electroporated with SRSF1 crRNA RNPs, whilst successfully activated, did not experience any loss of SRSF1 expression compared to negative control crRNA electroplated cells or non-electroporated cells at day 4 or day 6 post activation. This result was not unexpected; studies have methodically optimised CRISPR-Cas9 mediated knockouts in primary T cells over a plethora of experiments to gain electroporation conditions for optimal knockout efficiency (Integrated DNA Technologies, 2022) (Jordan, et al., 2008) (Seki & Rutz, 2018). It is unfortunate therefore, that we were only able to carry out one repeat of the experiment within the timeframe of the experiment. However, initial results in establishing the protocol for integration of electroporation mediated RNP transfection into *in vitro* CD4+ T cell activation assays lay a foundation for further optimisation to increase SRSF1, or other gene target, knockout efficiency in naïve CD4+ T cells. Alternatively, a possible route forward lies within the use of *Srsf1*-cKO mice (Katsuyama, et al., 2019) as a source of *Srsf1*-negative naïve CD4+ cells. Coupled with established and efficient *in vitro* gapmer knockdown of *Malat1* (Hewitson, et al., 2020), use of *SRSF1*-cKO derived naive CD4+ cells represents an equally valid alternative methodology of investigating the role of *Malat1* and SRSF1 in Th2 cell activation, compared to our own approach.

## 5.0 Conclusion

This project aimed to investigate the *Malat1* mediated riboregulation of SRSF1, specifically within Th2 polarised CD4+ T cell differentiation. We found that *Malat1* regulates RNA binding behaviour of SRSF1 in late stage differentiated Th2 cells, displaying the characterised riboregulatory axis between *Malat1* and SRSF1 in a novel context.

We specifically identified a role for *Malat1* in downregulating SRSF1-binding specifically to the distal isoform of *Runx3* in late stage differentiated Th2 cells. We subsequently showed

that *Malat1* upregulates *Runx3* and *dRunx3* within the same Th2 differentiation timepoints. These events require further investigation; specifically, within *Srsf1*$^{-/-}$ Th2 cells to confirm the role of SRSF1 in *Runx3* mRNA stability regulation.

In conclusion, our findings highlight one mechanism by which *Malat1* regulates Th2 differentiation, in the riboregulation of SRSF1-RNA interaction and the subsequent impact on expression of critical CD4+ T cell regulators across the Th2 differentiation trajectory.

# References


Korinfskaya, S., Parameswaran, S., Weirauch, M. . T. & Barski, A., 2021. Runx Transcription Factors in T Cells—What Is Beyond Thymic Development?. *Frontiers in Immunology.*

Abulwerdi, F. et al., 2019. Selective Small-Molecule Targeting of a Triple Helix Encoded by the Long Noncoding RNA, MALAT1. *ACS Chemical Biology,* pp. 223-235.

Agrelo, R. & Wutz, A., 2010. ConteXt of change—X inactivation and disease. *EMBO Molecular Medicine,* 2(1), pp. 6-15.

Arif, W. et al., 2023. Splicing factor SRSF1 deficiency in the liver triggers NASH-like pathology and cell death. *Nature Communications.*

Arun, G. et al., 2015. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes & Development,* Volume 30, pp. 34-51.

Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S., 2015. The MEME Suite. *Nucleic Acids Research,* 41(W1), pp. W39-W49.

Benard, D. et al., 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *The EMBO Journal,* Volume 29, pp. 3082-3093.

Bezrodnik, L. et al., 2013. Follicular bronchiolitis as phenotype associated with CD25 deficiency. *Clinical and Experimental Immunology,* Volume 175, pp. 227-234.

Botti, G. et al., 2019. LncRNA HOTAIR in Tumor Microenvironment: What Role?. *International Journal of Molecular Sciences,* 20(9).

Brannan, K. et al., 2021. Robust single-cell discovery of RNA targets of RNA binding proteins and ribosomes. *Nature Methods,* 18(5), pp. 507-519.

Brown, J. A. et al., 2012. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENβ noncoding RNAs. *PNAS,* Volume 109, p. 19202–19207.

Cáceres, J. et al., 1997. Role of the Modular Domains of SR Proteins in Subnuclear Localization and Alternative Splicing Specificity. *Journal of Cell Biology,* 138(2), p. 225–238.

Cano-Gamez, E. et al., 2020. Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nature Communications.*

Caudy, A. et al., 2007. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked-like syndrome, and defective IL-10 expression from CD4 lymphocytes. *Journal of Allergic and Clinical Immunology,* 119(2), pp. 482-487.

Cheng, Q. et al., 2022. CRISPR/Cas9 ribonucleoprotein (RNP) complex enables higher viability of transfected cells in genome editing of acute myeloid cells. *Annals of Translational Medicine,* 10(16).

Chen, Q., Zhu, C. & Jin, Y., 2020. The Oncogenic and Tumor Suppressive Functions of the Long Noncoding RNA MALAT1: An Emerging Controversy. *Frontiers in Genetics,* 11(93).

Chen, X. et al., 2020. Practical considerations on performing and analyzing CLIP-seq experiments to identify transcriptomic-wide RNA-Protein interactions. *Methods,* Volume 155, pp. 49-57.

Cléry, A. et al., 2021. Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon 7 splicing. *Nature Communications.*

Collier, S. P., Henderson, M. A., Tossberg, J. T. & Aune, T. M., 2014. Regulation of the Th1 genomic locus from Ifng through Tmevpg1 by T-bet. *Journal of Immunology,* 193(8), pp. 3959-65.

Derrien, T. et al., 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research,* 22(9), pp. 1775-1789.

Díaz-Muñoz, M. D. & Turner, M., 2018. Uncovering the Role of RNA-Binding Proteins in Gene Expression in the Immune System. *Fronteirs in Immunology,* Volume 9.

Du, J.-X.et al., 2021. SRSF1-Regulated Alternative Splicing in Breast Cancer Splicing factor SRSF1 promotes breast cancer progression via oncogenic splice switching of PTPMT1. *Journal of Experimental & Clinical Cancer Research.*

Egawa, T. et al., 2007. The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. *Journal of Experimental Medicine,* 204(8), pp. 1945-1957.

Eißmann, M. et al., 2012. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biology,* Volume 9, pp. 1076-1087.

El Said, H. N. et al., 2021. Malat-1-PRC2-EZH1 interaction supports adaptive oxidative stress dependent epigenome remodeling in skeletal myotubes. *Cell Death and Disease,* 12(850).

Engreitz, J. et al., 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell,* 159(1), pp. 188-199.

Fajrial, A. et al., 2020. A review of emerging physical transfection methods for CRISPR/Cas9-mediated gene editing. *Theranostics,* 10(12), pp. 5532-5549.

Flamand, M., Ke, K., Tamming, R. & Meyer, K., 2022. Single-molecule identification of the target RNAs of different RNA binding proteins simultaneously in cells. *Genes and Development,* Volume 36, pp. 1002-1015.

Gao, F., Tan, Y. & Luo, H., 2020. MALAT1 is involved in type I IFNs-mediated systemic lupus erythematosus by up-regulating OAS2, OAS3, and OASL. *Brazilian Journal of Medical and Biological Research,* 53(3).

Gibbons, H. R. et al., 2018. Divergent lncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells. *Fronteirs in Immunology.*

Gloria, V. G. et al., 2014. T Cell Activation Regulates CD6 Alternative Splicing by Transcription Dynamics and SRSF1. *The Journal of Immunology,* 193(1), pp. 391-399.

Gratacap, R., Jin, Y., Mantsopoulou, M. & Houston, R., 2020. Efficient Genome Editing in Multiple Salmonid Cell Lines Using Ribonucleoprotein Complexes. *Marine Biotechnology,* 22(5), pp. 717-724.

Gresch, O. et al., 2004. New non-viral method for gene transfer into primary cells. *Methods.*

Guo, M. et al., 2019. Structural insights into a high fidelity variant of SpCas9. *Cell Research,* pp. 183-192.

Gutschner, T. et al., 2013. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research,* p. 1180–1189.

Hafner, M. et al., 2021. CLIP and complementary methods. *Nature Reviews Methods Primers.*

Hafner, M. et al., 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell,* 141(1), pp. 129-141.

Hall, J. & Lekka, E., 2018. Noncoding RNAs in disease. *Febs Letters,* 592(17), p. 2884–2900.

Hall, L., Smith, K., Byron, M. & Lawrence, J., 2006. The Molecular Anatomy of a nuclear speckle. *Anat Rec A Discov Mol Cell Evol Biol,* 288(7), p. 664–675.

Haward, F. et al., 2021. Nucleo-cytoplasmic shuttling of splicing factor SRSF1 is required for development and cilia function. *eLife.*

Hedfors, I. A. & Brinchmann, J. E., 2003. Long-term proliferation and survival of in vitro-activated T cells is dependent on Interleukin-2 receptor signalling but not on the high-affinity IL-2R. *Scandinavian Journal of Immunology,* 58(5), pp. 522-532.

Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D., 2013. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell,* Volume 153, pp. 654-665.

Hentze, M., Castello, A., Schwarzl, T. & Preiss, T., 2018. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology,* p. 327–341.

Heubach, J. et al., 2015. The long noncoding RNA HOTAIR has tissue and cell type-dependent effects on HOX gene expression and phenotype of urothelial cancer cells. *Molecular Cancer,* 14(108).

Hewitson, J. P. et al., 2020. Malat1 Suppresses Immunity to Infection through Promoting Expression of Maf and IL-10 in Th Cells. *The Journal of Immunology,* 204(11), pp. 2949-2960.

Hori, S., Nomaru, T. & Sakaguchi, S., 2003. Control of regulatory T cell development by the transcription factor Foxp3. *Science.*

Howard, J. M. & Sanford, J. R., 2016. THE RNAissance Family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdisciplinary Reviews: RNA,* 6(1), pp. 93-110.

Huang, B., Guo, X. & Li, Y., 2020. lncRNA MALAT1 regulates the expression level of miR-21 and interferes with the biological behavior of colon cancer cells. *Journal of the Balkan Union of Oncology,* 25(2), pp. 907-913.

Hupertz, I. et al., 2014. iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods,* 65(3), pp. 274-287.

Hutchinson, J. N. et al., 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics,* 8(39).

Integrated DNA Technologies, 2022. Delivery of ribonucleoprotein complexes into Jurkat T cells using the Bio-Rad Gene Pulser® XcellTM Electroporation System.

Ip, J. Y. et al., 2007. Global analysis of alternative splicing during T-cell activation. *RNA,* Volume 13, p. 563–572.

Jäger, A. & Kuchroo, V. K., 2010. Effector and regulatory T cell subsets in autoimmunity and tissue inflammation. *Scandinavian Journal of Immunology,* 72(3), pp. 173-184.

Jensen, T. et al., 2021. Targeted regulation of transcription in primary cells using CRISPRa and CRISPRi. *Genome Research,* 31(11), pp. 2120-2130.

Ji, P. et al., 2003. MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene,* Volume 22, pp. 8031-8041.

Ji, Q. et al., 2019. MALAT1 regulates the transcriptional and translational levels of proto-oncogene RUNX2 in colorectal cancer metastasis. *Cell Death & Disease.*

Jobbins, A. M. et al., 2021. Exon-independent recruitment of SRSF1 is mediated by U1 snRNP stem-loop 3. *The EMBO Journal,* Volume 41.

Joosse, M., Charbit-Henrion, F., Boisgard, R. & Raatgeep, R., 2021. Duplication of the IL2RA locus causes excessive IL-2 signaling and may predispose to very early onset colitis. *Nature Mucosal Immunology,* Volume 14, p. 1172–1182.

Jordan, E. et al., 2008. Optimizing Electroporation Conditions in Primary and Other Difficult-to-Transfect Cells. *Journal of Biomedical Technology,* 19(5), pp. 328-334.

Ju, A. et al., 2019. A carrier-free multiplexed gene editing system applicable for suspension cells. *Biomaterials.*

Juarez, I. et al., 2022. Splicing factor SRSF1 is essential for CD8 T cell function and host antigen-specific viral immunity. *Frontier in Immunology,* Volume 13.

Katsuyama, T. et al., 2019. Splicing factor SRSF1 controls T cell hyperactivity and systemic autoimmunity. *The Journal of Clinical Investigation,* 129(12), pp. 5411-5423.

Katsuyama, T. & Moulton, V. R., 2021. Splicing factor SRSF1 is indispensable for regulatory T cell homeostasis and function. *Cell Reports.*

Kim, B., Sasak, Y. & Egawa, T., 2015. Restriction of non-permissive RUNX3 protein expression in T lymphocytes by the Kozak sequence. *Journal of Immunology,* 195(4), p. 1517–1523.

Kim, J. et al., 2018. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nature Genetics,* Volume 50, p. s1705–1715.

Ko, H., Kim, C. J. & Im, S.-H., 2022. T Helper 2-Associated Immunity in the Pathogenesis of Systemic Lupus Erythematosus. *Frontiers in Immunology,* Volume 13.

Kohtz, J. D. et al., 1994. Protein-protein interactions and 5' -splice-site recognition in mammalian mRNA precursors. *Nature,* pp. 119-124.

Kohu, K. et al., 2009. The Runx3 Transcription Factor Augments Th1 and Down-Modulates Th2 Phenotypes by Interacting with and Attenuating GATA3. *The Journal of Immunology,* p. 7817–7824.

Kono, M. et al., 2018. Decreased Expression of Serine/Arginine-Rich Splicing Factor 1 in T Cells From Patients With Active Systemic Lupus Erythematosus Accounts for Reduced Expression of RasGRP1 and DNA Methyltransferase 1. *Arthritis & Rheumatology,* 70(12), pp. 2046-2056.

Krichevsky, A. & Gabriely, G., 2009. miR-21: a small multi-faceted RNA. *Journal of Cellular and Molecular Medicine,* 13(1), pp. 39-53.

Kubo, M., 2017. Innate and adaptive type 2 immunity in lung allergic inflammation. *Immunology Review,* pp. 162-172.

Langrish, C. et al., 2005. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *Journal of Experimental Medicine,* pp. 233-240.

Liang, X. et al., 2015. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology,* pp. 44-53.

Liang, X.-H., Sun, H., Nichols, J. & Crooke, S., 2017. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Molecular Therapy,* 25(9), pp. 2017-2092.

Liao, S. & Regev, O., 2021. Splicing at the phase-separated nuclear speckle interface: a model. *Nucleic Acids Research,* 49(2), p. 636–645.

Li, J. et al., 2018. Clinicopathological and prognostic significance of long noncoding RNA MALAT1 in human cancers a review and meta-analysis. *Cancer Cell International,* 18(109).

Littman, D. R. & Egawa, T., 2008. ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. *Nature Immunology,* 9(10), pp. 1131-1139.

Liu, C., Zhang, Y., Ma, Z. & Yi, H., 2022. Long Noncoding RNAs as Orchestrators of CD4+ T-Cell Fate. *Frontiers in Cell and Developmental Biology,* Volume 10.

Liu, H.-X., Zhang, M. & Krainer, A., 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes and Development,* 12(13), pp. 1998-2011.

Li, X. & Manley, J. L., 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell,* 122(3), pp. 365-378.

Lloyd, C. M. & Snelgrove, R. J., 2018. Type 2 immunity: Expanding our view. *Science Immunology,* Volume 3.

Loda, A. & Heard, E., 2019. Xist RNA in action: Past, present, and future. *PLOS Genetics,* 15(9).

Luckheeram, R., Zhou, R., Verma, A. & Xia, B., 2012. CD4+T Cells: Differentiation and Functions. *Clinical and Developmental Immunology,* pp. 1-12.

Lynch, K. & Motta-Mena, L., 2013. Alternative Splicing. *Encyclopedia of Biological Chemistry,* pp. 75-80.

Maier, L. et al., 2009. Soluble IL-2RA Levels in Multiple Sclerosis Subjects and the Effect of Soluble IL-2RA on Immune Responses1. *Journal of Immunology,* 182(3), pp. 1541-1547.

Malakar, P. et al., 2019. Long Noncoding RNA MALAT1 Regulates Cancer Glucose Metabolism by Enhancing mTOR-Mediated Translation of TCF7L2. *Cancer Research: Metabolism and Chemical Biology,* 79(10), pp. 2480-2493.

Ma, L., Bajic, V. & Zhang, Z., 2013. On the classification of long non-coding RNAs. *RNA biology,* 10(6), pp. 924-933.

Marahens, Y. et al., 1997. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev,* 11(2), pp. 155-156.

Maslon, M. et al., 2014. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife,* pp. 1-27.

Masoumi, F. et al., 2019. Malat1 long noncoding RNA regulates inflammation and leukocyte differentiation in experimental autoimmune encephalomyelitis. *Journal of Neuroimmunology,* 15(328), pp. 50-59.

Matera, A. G., Terns, R. M. & Terns, M. P., 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology,* Volume 8, pp. 209-220.

McGeachy, M. & Cua, D., 2008. Th17 Cell Differentiation: The Long and Winding Road. *Immunity,* pp. 445-453.

McHugh, C. A. & Guttman, M., 2018. RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells. *Methods of Molecular Biology,* pp. 473-488.

McMahon, A. et al., 2016. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell,* Volume 165, pp. 742-753.

Minami, Y., Kono, T., Miyazaki, T. & Taniguchi, T., 1993. The IL2 Receptor Complex: its structure, function and target genes. *Annual Review of Immunology,* Issue 11, pp. 245-268.

Miyamoto, C. et al., 2019. Runx/Cbfβ complexes protect group 2 innate lymphoid cells from exhausted-like hyporesponsiveness during allergic airway inflammation. *Nature Communications.*

Moore, P. & Steitz, T., 2011. The Roles of RNA in the Synthesis of Protein. *CSB Perspectives in Biology.*

Morais, P., Adachi, H. & Yu, Y.-T., 2021. Spliceosomal snRNA Epitranscriptomics. *Frontiers in Genetics,* Volume 12.

Motta-Mena, L., Heyd, F. & Lynch, K., 2010. Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Molecular Cell,* 37(2).

Moulton, V., Grammatikos, A., Fitzgerald, L. & Tsokos, G., 2013. Splicing factor SF2/ASF rescues IL-2 production in T cells from systemic lupus erythematosus patients by activating IL-2 transcription. *PNAS,* 110(5), p. 1845–1850.

Nakagawa, S. et al., 2012. Malat1 is not an essential component of nuclear speckles in mice. *RNA,* Volume 18, p. 1487–1499.

Naoe, Y. et al., 2007. Repression of interleukin-4 in T helper type 1 cells by Runx/Cbfβ binding to the Il4 silencer. *Journal of Experimental Medicine,* pp. 1749-1755.

Napotnik, T., Polajzer, T. & Miklavcic, D., 2021. Cell death due to electroporation – A review. *Bioelectrochemistry.*

O'Brien, J., Hayder, H., Zayed, Y. & Peng, C., 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology,* 9(402).

Oh, S. et al., 2022. High-efficiency nonviral CRISPR/Cas9-mediated gene editing of human T cells using plasmid donor DNA. *Journal of Experimental Medicine,* 219(5).

Paz, S., Ritchie, A., Mauer, C. & Caputi, M., 2021. The RNA binding protein SRSF1 is a master switch of gene expression and regulation in the immune system. *Cytokine Growth Factor Review,* Issue 57, p. 19–26.

Plasek, L. M. & Valadkhan, S., 2021. lncRNAs in T lymphocytes: RNA regulation at the heart of the immune response. *American Journal Physiological, Cell Physiology,* pp. 415-427.

Qi, Z. et al., 2021. SRSF1 serves as a critical posttranscriptional regulator at the late stage of thymocyte development. *Immunology,* Volume 7, pp. 1-17.

Quattrone, A. & Dassi, E., 2019. The Architecture of the Human RNA-Binding Protein Regulatory Network. *iScience,* pp. 706-719.

Qu, D. et al., 2019. Long noncoding RNA MALAT1 releases epigenetic silencing of HIV-1 replication by displacing the polycomb repressive complex 2 from binding to the LTR promoter. *Nucleic Acids Research,* 47(6), pp. 3013-3027.

Rahimmanesh, I., Totonchi, M., Khanahmad, H. & Rahimmanesh, I., 2020. The challenging nature of primary T lymphocytes for transfection: Effect of protamine sulfate on the transfection efficiency of chemical transfection reagents. *Research in Pharmaceutical Sciences,* 15(5), pp. 437-446.

Ran , F. A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols,* Volume 8, p. 2281–2308.

Ran, F. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocol,* pp. 2281-2308.

Ranzani, V. et al., 2015. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nature Immunology,* Volume 16, p. 318–325.

Ratnadiwakara, M. & Änkö, M.-L., 2018. mRNA Stability Assay Using transcription inhibition by Actinomycin D in Mouse Pluripotent Stem Cells. *Biology Protocol,* 8(21).

Rennick, D. M., Fort, M. M. & Davidson, N. J., 2019. Studies with IL-10-/- mice: an overview. *Journal of Leukocyte Biology,* 61(4), p. 389–396.

Ross, S. & Cantrell, D., 2018. Signaling and Function of Interleukin-2 in T Lymphocytes. *Annunal Review of Immunology,* Issue 36, pp. 411-433.

Salmena, L. et al., 2011. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language?. *Cell,* 146(3), pp. 353-358.

Sanford, J. R. et al., 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Research,* pp. 381-394.

Saraiva, M. & O'Garra, A., 2010. The regulation of IL-10 production by immune cells. *Nature Reviews Immunology,* Volume 10, p. 170–181.

Saumik, . B. et al., 2018. MALAT1: An Epigenetic Regulator of Inflammation in Diabetic Retinopathy. *Nature Scientific Reports,* Volume 8, p. 6526.

Scherer, M., Levin, M., Butter, F. & Scheibe, M., 2020. Quantitative Proteomics to Identify Nuclear RNA-Binding Proteins of Malat1. *International Journal of Molecular Sciences,* 21(3).

Seki, A. & Rutz, S., 2018. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockouts in Primary T cells. *Journal of Experimental Medicine,* 215(3), p. 985–997.

Shepard, P. j. & Hertel, K. J., 2009. The SR protein family. *Genome Biology,* Volume 242.

Spector, D. L. & Lamond, A. I., 2011. Nuclear Speckles. *CSH Perspectives in Biology,* 3(2).

Spellberg, B. & Edwards Jr, J., 2001. Type 1/Type 2 Immunity in Infectious Diseases. *Clinical Infectious Diseases,* 32(1), pp. 76-102.

Spiniello, M. et al., 2018. HyPR-MS for Multiplexed Discovery of MALAT1, NEAT1, and NORAD lncRNA Protein Interactomes. *Journal of Proteome Research,* 17(9), p. 3022–3038.

Stamm, S., 2008. Regulation of Alternative Splicing by Reversible Protein Phosphorylation. *Journal of Biological Chemistry,* 283(3), pp. 1223-1227.

Stauber, D. et al., 2006. Crystal structure of the IL-2 signaling complex: Paradigm for a heterotrimeric cytokine receptor. *PNAS,* 103(8), pp. 2788-2793.

Stein, C. et al., 2010. Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic Acids Research,* 38(1).

Su, K. et al., 2021. The role of a ceRNA regulatory network based on lncRNA MALAT1 site in cancer progression. *Biomedicine and Pharmacotherapy,* Volume 137.

Su, S. et al., 2016. CRISPR-Cas9 mediated efficient PD-1 disruption on human primary T cells from cancer patients. *Nature Scientific Reports,* Volume 6.

Szabo, S. et al., 2000. A Novel Transcription Factor, T-bet, Directs Th1 Lineage Commitment. *Cell,* 100(6), pp. 655-669.

Tai, Y. et al., 2019. Dysregulation of humoral immunity in Foxp3 conditional-knockout mice. *Biochemical and Biophysical Research Communications,* pp. 787-793.

Tan, J. T. et al., 2001. IL-7 is critical for homeostatic proliferation and survival of naïve T cells. *PNAS,* 98(15), pp. 8732-8737.

Techasintana, P. et al., 2017. The RNA-Binding Protein HuR Posttranscriptionally Regulates IL-2 Homeostasis and CD4+ Th2 Differentiation. *Immunohorizons,* 1(6), pp. 109-123.

The ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* Volume 489, p. 57–74.

Tripathi, V. et al., 2010. The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Molecular Cell,* 39(6), p. 925–938.

Wang, J., Smith, P., Krainer, A. & Zhang, M., 2005. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic acids research,* 33(16), p. 5053–5062.

Wang, X. et al., 2016. LncRNA MALAT1 promotes development of mantle cell lymphoma by associating with EZH2. *Journal of Translational Medicine,* 20(14).

West, J. et al., 2014. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Molecular Cell,* 55(5), pp. 791-802.

West, K. A. & Lagos, D., 2019. Long Non-Coding RNA Function in CD4+ T cells: What We Know and What Next. *non-coding RNA,* 5(43).

Wightman, B., Ha, I. & Ruvkun, G., 1993. Posttranscriptional Regulation of the Heterochronic Gene lin-14 by W-4 Mediates Temporal Pattern Formation in C. elegans. *Cell,* 75(5), pp. 855-862.

Wilusz, J. E. et al., 2012. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *GENES & DEVELOPMENT 26:2392–2407,* Volume 26, pp. 2392-2407.

Wong, L., Brampton, C., Woo, D. & Dresken, E., 2019. Optimizing Electroporation Conditions for High-Efficiency mRNA Transfection of CD8+ T Cells with the Gene Pulser Xcell Electroporation System. *Bio-Rad.*

Wong, W. F. et al., 2011. Interplay of transcription factors in T-cell differentiation and function: the role of Runx. *Immunology,* Issue 132, p. 157–164.

Xu, X. et al., 2018. Efcient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tub electroporation. *Scientific Reports.*

Yang, H. et al., 2017. Long noncoding RNA MALAT-1 is a novel inflammatory regulator in human systemic lupus erythematosus. *Oncotarget,* 8(44), pp. 77400-77406.

Yang, T. et al., 2017. Upregulation of Bcl-2 and Its Promoter Signals in CD4+ T Cells during Neuromyelitis Optica Remission. *Frontiers in Neuroscience1,* 11(11).

Yang, X. et al., 2008. T Helper 17 Lineage Differentiation Is Programmed by Orphan Nuclear Receptors RORalpha and RORgamma.. *Immunity,* pp. 29-39.

Yang, Y.-C.T. et al., 2015. CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics,* Volume 51.

Yan, Y. et al., 2012. TCR stimulation upregulates MS4a4B expression through induction of AP-1 transcription factor during T cell activation. *Molecular Immunology,* 52(2), pp. 71-78.

Yao, R., Wang , Y. & Chen, L., 2019. Cellular Function of long noncoding RNAs. *Nature Cell Biology,* Volume 21, pp. 542-551.

Yao, Y. et al., 2018. Long noncoding RNA Malat1 is not essential for T cell development and response to LCMV infection. *RNA BIOLOGY,* 15(12), p. 1477–1486.

Zetsche, B. et al., 2017. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array.. *Nature Biotechnology,* 35(1), pp. 31-34.

Zhang, B. et al., 2012. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep,* Volume 2, pp. 111-123.

Zhang, X., Hamblin, M. & Yin, K.-J., 2017. The long noncoding RNA Malat1: Its physiological and pathophysiological functions. 14(12), pp. 1705-1714.

Zhang, Y. et al., 2019. LncRNA MALAT1 promotes osteoarthritis by modulating miR-150-5p/AKT3 axis. *Cell and Bioscience,* 9(54).

Zhao, Y. et al., 2016. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Research,* Volume 44, pp. 203-209.

Zhou, J., Alvarez, C. & Cobb, B., 2021. Integration of IL-2 and IL-4 signals coordinates divergent regulatory T cell responses and drives therapeutic efficacy. *eLife.*

Zhu, J. et al., 2006. GATA-3 promotes Th2 responses through three different mechanisms: induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of Th1 cell-specific factors. *Cell Research,* Volume 16, pp. 3-10.

# Appendix

## Appendix 1.

Details of Transcripts with a greater than 1.5-fold increase/decrease in $Log_2$(csFC) or abundance normalised $Log_2$(csFC).

| Gene Name | Ensembl ID | $Log_2$(SRSF1 crosslink signal fold change) | $Log_2$(Abundance fold change) | Normalised $Log_2$(SRSF1 crosslink signal fold change) |
|---|---|---|---|---|
| Ppp2r3a | ENSMUSG00000043154.15 | -2.74 | -0.27 | -2.48 |
| Adamtsl3 | ENSMUSG00000070469.12 | -2.62 | -1.52 | -1.10 |
| Mctp1 | ENSMUSG00000021596.16 | -2.59 | 0.77 | -3.36 |
| Egln3 | ENSMUSG00000035105.5 | -2.15 | -1.42 | -0.72 |
| P4ha1 | ENSMUSG00000019916.14 | -1.79 | -0.35 | -1.44 |
| Cysltr1 | ENSMUSG00000052821.3 | -1.77 | -1.41 | -0.36 |
| Gm20275 | ENSMUSG00000110803.1 | -1.70 | 0.28 | -1.98 |
| mt-Rnr2 | ENSMUSG00000064339.1 | 1.62 | -0.18 | 1.80 |
| Actn1 | ENSMUSG00000015143.15 | 1.53 | 1.08 | 0.45 |
| Patj | ENSMUSG00000061859.16 | 1.50 | 0.42 | 1.08 |
| Irgm2 | ENSMUSG00000069874.7 | 1.49 | 0.06 | 1.43 |
| mt-Nd5 | ENSMUSG00000064367.1 | 1.49 | -0.12 | 1.61 |
| Chn2 | ENSMUSG00000004633.17 | -1.49 | -0.47 | -1.02 |
| Btla | ENSMUSG00000052013.14 | -1.47 | -0.07 | -1.40 |
| Pip5k1b | ENSMUSG00000024867.14 | -1.42 | -0.22 | -1.20 |
| Olfr60 | ENSMUSG00000060112.4 | -1.41 | -0.97 | -0.44 |
| A430093F15Rik | ENSMUSG00000067577.8 | -1.41 | -0.06 | -1.34 |

| | | | | |
|---|---|---|---|---|
| Ero1l | ENSMUSG00000021831.9 | -1.40 | -1.24 | -0.16 |
| Gata3 | ENSMUSG00000015619.10 | -1.39 | -0.23 | -1.16 |
| Gramd3 | ENSMUSG00000001700.9 | 1.36 | 1.29 | 0.08 |
| Mboat1 | ENSMUSG00000038732.15 | -1.35 | -0.30 | -1.05 |
| Nebl | ENSMUSG00000053702.16 | -1.31 | 0.74 | -2.05 |
| Zhx2 | ENSMUSG00000071757.10 | 1.30 | 0.77 | 0.53 |
| Mctp2 | ENSMUSG00000032776.9 | 1.28 | -0.02 | 1.30 |
| Cdc25b | ENSMUSG00000027330.16 | -1.28 | 0.00 | -1.28 |
| Sntb1 | ENSMUSG00000060429.12 | 1.25 | 1.61 | -0.36 |
| Stat1 | ENSMUSG00000026104.14 | 1.25 | -0.20 | 1.45 |
| Mxi1 | ENSMUSG00000025025.13 | -1.24 | -0.70 | -0.54 |
| Gpi1 | ENSMUSG00000036427.5 | -1.18 | -0.32 | -0.86 |
| Ccr2 | ENSMUSG00000049103.14 | -1.18 | 0.70 | -1.88 |
| mt-Rnr1 | ENSMUSG00000064337.1 | 1.14 | 0.16 | 0.98 |
| Cpq | ENSMUSG00000039007.10 | -1.11 | 0.32 | -1.43 |
| Vim | ENSMUSG00000026728.9 | -1.11 | -0.96 | -0.15 |
| Pkm | ENSMUSG00000032294.17 | -1.11 | -0.76 | -0.35 |
| Maml3 | ENSMUSG00000061143.15 | 1.10 | -0.56 | 1.66 |
| Chst11 | ENSMUSG00000034612.7 | 1.08 | -0.75 | 1.83 |
| B4galnt1 | ENSMUSG00000006731.10 | 1.07 | 0.28 | 0.79 |
| Mki67 | ENSMUSG00000031004.8 | -1.07 | -0.50 | -0.57 |
| Itpr1 | ENSMUSG00000030102.11 | -1.06 | -0.28 | -0.78 |

| | | | | |
|---|---|---|---|---|
| Plcl1 | ENSMUSG00000038349.10 | -1.04 | 0.00 | -1.03 |
| Snx25 | ENSMUSG00000038291.16 | -1.03 | 0.00 | -1.03 |
| Neat1 | ENSMUSG00000092274.2 | -1.03 | -0.56 | -0.47 |
| Diaph3 | ENSMUSG00000022021.14 | -1.02 | -0.51 | -0.51 |
| Cracr2a | ENSMUSG00000061414.8 | 1.02 | 0.93 | 0.09 |
| Epsti1 | ENSMUSG00000022014.15 | 1.00 | 0.04 | 0.97 |
| Tespa1 | ENSMUSG00000034833.10 | -1.00 | 0.06 | -1.06 |
| Ctla4 | ENSMUSG00000026011.13 | -0.99 | -0.68 | -0.30 |
| Gm4759 | ENSMUSG00000053541.3 | -0.97 | 0.99 | -1.96 |
| Top2a | ENSMUSG00000020914.17 | -0.96 | -0.49 | -0.48 |
| Slc20a1 | ENSMUSG00000027397.14 | 0.96 | 0.67 | 0.29 |
| Pgm2 | ENSMUSG00000025791.18 | -0.95 | -0.34 | -0.61 |
| Acsl4 | ENSMUSG00000031278.12 | -0.94 | -0.09 | -0.85 |
| Xist | ENSMUSG00000086503.3 | -0.93 | -0.03 | -0.90 |
| Parp8 | ENSMUSG00000021725.9 | 0.92 | 0.88 | 0.04 |
| Ccr5 | ENSMUSG00000079227.10 | 0.92 | -1.47 | 2.39 |
| Pfkp | ENSMUSG00000021196.14 | -0.92 | -0.53 | -0.39 |
| Rps6ka5 | ENSMUSG00000021180.9 | -0.92 | -0.41 | -0.51 |
| Rapgef4 | ENSMUSG00000049044.16 | 0.90 | 0.72 | 0.18 |
| Clcn3 | ENSMUSG00000004319.15 | -0.89 | -0.50 | -0.40 |
| Smyd3 | ENSMUSG00000055067.15 | 0.88 | 0.17 | 0.71 |
| Pim1 | ENSMUSG00000024014.7 | -0.87 | -0.96 | 0.09 |

| | | | | |
|---|---|---|---|---|
| Il2ra | ENSMUSG00000026770.5 | -0.87 | -1.33 | 0.46 |
| Syne3 | ENSMUSG00000054150.12 | -0.87 | -0.13 | -0.73 |
| Ms4a4b | ENSMUSG00000056290.15 | 0.86 | 1.72 | -0.86 |
| Ahnak | ENSMUSG00000069833.12 | -0.86 | 0.44 | -1.29 |
| Aff4 | ENSMUSG00000049470.13 | -0.85 | -0.47 | -0.38 |
| Igf1r | ENSMUSG00000005533.10 | 0.85 | 0.15 | 0.70 |
| Zfp292 | ENSMUSG00000039967.14 | -0.84 | -0.08 | -0.76 |
| Klhl3 | ENSMUSG00000014164.15 | 0.84 | 1.05 | -0.21 |
| Bbs9 | ENSMUSG00000035919.16 | -0.83 | 0.38 | -1.20 |
| Pag1 | ENSMUSG00000027508.15 | -0.82 | 0.41 | -1.23 |
| St8sia4 | ENSMUSG00000040710.10 | -0.82 | 0.01 | -0.83 |
| Cep170 | ENSMUSG00000057335.11 | -0.81 | -0.68 | -0.13 |
| Kdm3a | ENSMUSG00000053470.13 | -0.81 | -0.26 | -0.54 |
| Npepps | ENSMUSG00000001441.13 | -0.80 | -0.30 | -0.51 |
| Map4k4 | ENSMUSG00000026074.14 | -0.79 | 0.36 | -1.15 |
| Ly6e | ENSMUSG00000022587.14 | 0.79 | 0.30 | 0.48 |
| Ipcef1 | ENSMUSG00000064065.15 | 0.78 | 1.19 | -0.41 |
| Slco3a1 | ENSMUSG00000025790.14 | 0.78 | -0.05 | 0.83 |
| Ttc7b | ENSMUSG00000033530.8 | 0.78 | 0.01 | 0.77 |
| Gm8995 | ENSMUSG00000063286.7 | -0.78 | 0.08 | -0.85 |
| Mipol1 | ENSMUSG00000047022.18 | 0.77 | 0.18 | 0.59 |
| Rap1gap2 | ENSMUSG00000038807.18 | -0.77 | 1.17 | -1.95 |

| | | | | |
|---|---|---|---|---|
| Adk | ENSMUSG00000039197.10 | -0.77 | -0.03 | -0.74 |
| Cmss1 | ENSMUSG00000022748.7 | 0.77 | -0.45 | 1.22 |
| Rbpj | ENSMUSG00000039191.12 | -0.77 | -0.70 | -0.07 |
| Themis | ENSMUSG00000049109.15 | -0.75 | -0.01 | -0.74 |
| Gpm6b | ENSMUSG00000031342.17 | -0.75 | -1.47 | 0.72 |
| Serinc3 | ENSMUSG00000017707.9 | -0.75 | -0.17 | -0.58 |
| Ezh2 | ENSMUSG00000029687.16 | -0.75 | -0.54 | -0.20 |
| Itgb3 | ENSMUSG00000020689.4 | -0.75 | 0.27 | -1.02 |
| Sorl1 | ENSMUSG00000049313.8 | 0.74 | 0.43 | 0.31 |
| Mbd5 | ENSMUSG00000036792.12 | -0.73 | 0.19 | -0.93 |
| Samd9l | ENSMUSG00000047735.14 | -0.73 | 0.57 | -1.30 |
| Pde7a | ENSMUSG00000069094.12 | 0.73 | 0.34 | 0.39 |
| Rap1gds1 | ENSMUSG00000028149.12 | -0.73 | -0.29 | -0.44 |
| Abcg3 | ENSMUSG00000029299.14 | -0.73 | 0.00 | -0.73 |
| Tmlhe | ENSMUSG00000079834.2 | -0.72 | 0.15 | -0.87 |
| Samsn1 | ENSMUSG00000022876.18 | -0.72 | -0.46 | -0.26 |
| Nfia | ENSMUSG00000028565.18 | -0.71 | -0.41 | -0.30 |
| Slc4a7 | ENSMUSG00000021733.10 | -0.71 | 0.02 | -0.74 |
| Traf1 | ENSMUSG00000026875.14 | 0.71 | 0.24 | 0.47 |
| Pigk | ENSMUSG00000039047.17 | 0.71 | -0.07 | 0.78 |
| Crybg1 | ENSMUSG00000019866.14 | -0.70 | -0.53 | -0.17 |
| Sidt1 | ENSMUSG00000022696.17 | 0.70 | 0.91 | -0.21 |

| | | | | |
|---|---|---|---|---|
| Jak2 | ENSMUSG00000024789.12 | -0.69 | -0.19 | -0.50 |
| Icos | ENSMUSG00000026009.14 | -0.69 | -0.75 | 0.05 |
| Ar | ENSMUSG00000046532.8 | -0.68 | 0.46 | -1.15 |
| Cd44 | ENSMUSG00000005087.17 | -0.68 | -0.31 | -0.37 |
| Zfp942 | ENSMUSG00000071267.11 | -0.68 | 0.38 | -1.06 |
| Tle4 | ENSMUSG00000024642.16 | 0.68 | 1.08 | -0.40 |
| Rlf | ENSMUSG00000049878.13 | -0.67 | 0.02 | -0.69 |
| Trp53inp1 | ENSMUSG00000028211.11 | -0.67 | 0.20 | -0.87 |
| Ifngr1 | ENSMUSG00000020009.12 | 0.67 | -0.25 | 0.92 |
| Runx3 | ENSMUSG00000070691.10 | 0.66 | -0.77 | 1.43 |
| Tcf7 | ENSMUSG00000000782.15 | 0.66 | 0.92 | -0.25 |
| Zfp654 | ENSMUSG00000047141.5 | -0.66 | 0.17 | -0.83 |
| Foxn3 | ENSMUSG00000033713.12 | -0.66 | 0.01 | -0.67 |
| Tgfbr1 | ENSMUSG00000007613.15 | -0.66 | -0.03 | -0.63 |
| Tk2 | ENSMUSG00000035824.7 | 0.66 | 0.59 | 0.07 |
| Rgs3 | ENSMUSG00000059810.18 | 0.65 | 0.25 | 0.40 |
| Alcam | ENSMUSG00000022636.13 | -0.64 | -0.77 | 0.13 |
| Chm | ENSMUSG00000025531.14 | -0.64 | -0.57 | -0.06 |
| Gm1966 | ENSMUSG00000073902.5 | -0.64 | 0.33 | -0.97 |
| Gpr146 | ENSMUSG00000044197.8 | -0.63 | -0.07 | -0.56 |
| Plgrkt | ENSMUSG00000016495.12 | 0.63 | 0.21 | 0.42 |
| Arrb1 | ENSMUSG00000018909.15 | -0.63 | -0.31 | -0.31 |

| | | | | |
|---|---|---|---|---|
| Atp10d | ENSMUSG00000046808.17 | -0.61 | 0.55 | -1.16 |
| Tet2 | ENSMUSG00000040943.12 | -0.61 | -0.22 | -0.39 |
| Cd5 | ENSMUSG00000024669.7 | -0.61 | 0.34 | -0.96 |
| Tlk1 | ENSMUSG00000041997.16 | -0.61 | 0.30 | -0.91 |
| Nrip1 | ENSMUSG00000048490.13 | -0.61 | -0.15 | -0.46 |
| Camkmt | ENSMUSG00000071037.5 | 0.60 | 0.60 | 0.00 |
| Rexo5 | ENSMUSG00000030924.16 | -0.59 | 0.24 | -0.83 |
| Fgf13 | ENSMUSG00000031137.17 | 0.59 | -0.04 | 0.63 |
| Gbp4 | ENSMUSG00000079363.7 | 0.58 | -0.09 | 0.67 |
| Spn | ENSMUSG00000051457.7 | -0.58 | 0.14 | -0.72 |
| Slc17a9 | ENSMUSG00000023393.15 | -0.57 | 0.46 | -1.03 |
| Tmcc1 | ENSMUSG00000030126.17 | -0.57 | 0.09 | -0.66 |
| Txnip | ENSMUSG00000038393.14 | -0.56 | 0.33 | -0.89 |
| Atp11b | ENSMUSG00000037400.17 | -0.55 | 0.14 | -0.70 |
| Xpo6 | ENSMUSG00000000131.15 | -0.55 | 0.18 | -0.73 |
| Kcnq1ot1 | ENSMUSG00000101609.1 | -0.54 | 0.17 | -0.70 |
| Nlrc5 | ENSMUSG00000074151.13 | 0.53 | -0.19 | 0.72 |
| Cnot6l | ENSMUSG00000034724.17 | -0.53 | 0.13 | -0.66 |
| Ccdc82 | ENSMUSG00000079084.10 | -0.52 | 0.26 | -0.78 |
| Sacs | ENSMUSG00000048279.18 | -0.51 | 0.17 | -0.67 |
| Vav3 | ENSMUSG00000033721.16 | -0.50 | 0.51 | -1.02 |
| Fry | ENSMUSG00000056602.11 | -0.50 | 0.56 | -1.06 |

| | | | | |
|---|---|---|---|---|
| S1pr1 | ENSMUSG00000045092.8 | -0.50 | 0.49 | -0.99 |
| Cd96 | ENSMUSG00000022657.9 | 0.50 | 1.30 | -0.80 |
| Hmgb1-ps8 | ENSMUSG00000097295.1 | -0.50 | 1.44 | -1.93 |
| 4933406I18Rik | ENSMUSG00000087475.3 | -0.49 | 0.69 | -1.18 |
| Rap1b | ENSMUSG00000052681.8 | -0.49 | 0.12 | -0.61 |
| Rcsd1 | ENSMUSG00000040723.14 | -0.48 | 0.33 | -0.81 |
| Adnp | ENSMUSG00000051149.15 | -0.46 | 0.24 | -0.70 |
| Hexb | ENSMUSG00000021665.9 | 0.46 | -0.27 | 0.73 |
| Zc3h12d | ENSMUSG00000039981.6 | 0.45 | -0.38 | 0.83 |
| Nlrc3 | ENSMUSG00000049871.13 | -0.45 | 0.48 | -0.93 |
| Akt3 | ENSMUSG00000019699.16 | -0.45 | 0.22 | -0.67 |
| 1810026B05Rik | ENSMUSG00000101970.6 | -0.44 | 0.20 | -0.64 |
| Ank | ENSMUSG00000022265.7 | -0.44 | 0.33 | -0.77 |
| Atrx | ENSMUSG00000031229.16 | -0.43 | 0.21 | -0.64 |
| Ptpn4 | ENSMUSG00000026384.13 | -0.43 | 0.17 | -0.60 |
| Dym | ENSMUSG00000035765.9 | 0.43 | -0.32 | 0.75 |
| Dnah8 | ENSMUSG00000033826.9 | -0.42 | 0.42 | -0.84 |
| Ankrd12 | ENSMUSG00000034647.14 | -0.42 | 0.44 | -0.86 |
| Tcf20 | ENSMUSG00000041852.14 | -0.41 | 0.26 | -0.67 |
| Gm26917 | ENSMUSG00000097971.3 | -0.40 | 0.26 | -0.66 |
| Rasgrp2 | ENSMUSG00000032946.16 | -0.40 | 0.53 | -0.92 |
| Slc28a2 | ENSMUSG00000027219.13 | -0.39 | 0.38 | -0.78 |

| | | | | |
|---|---|---|---|---|
| Stap1 | ENSMUSG00000029254.16 | -0.39 | 0.41 | -0.80 |
| Tbc1d22a | ENSMUSG00000051864.9 | 0.39 | -0.32 | 0.71 |
| 4930523C07Rik | ENSMUSG00000090394.8 | -0.39 | 0.54 | -0.93 |
| Mdn1 | ENSMUSG00000058006.12 | 0.38 | 1.17 | -0.79 |
| Abcc4 | ENSMUSG00000032849.14 | -0.37 | 0.70 | -1.07 |
| Tet3 | ENSMUSG00000034832.15 | -0.37 | 0.46 | -0.82 |
| Sox5 | ENSMUSG00000041540.16 | -0.35 | -1.00 | 0.66 |
| Fnbp1 | ENSMUSG00000075415.13 | 0.35 | -0.28 | 0.62 |
| Flna | ENSMUSG00000031328.15 | -0.34 | 0.27 | -0.61 |
| Pdcd4 | ENSMUSG00000024975.12 | -0.34 | 0.53 | -0.87 |
| Macf1 | ENSMUSG00000028649.18 | -0.34 | 0.74 | -1.08 |
| Cdk6 | ENSMUSG00000040274.11 | -0.33 | -1.26 | 0.92 |
| Ccnd3 | ENSMUSG00000034165.16 | 0.33 | -0.51 | 0.84 |
| B630019A10Rik | ENSMUSG00000068463.3 | -0.33 | 1.65 | -1.98 |
| Nabp1 | ENSMUSG00000026107.11 | 0.33 | -0.65 | 0.98 |
| Sell | ENSMUSG00000026581.14 | -0.32 | 0.61 | -0.93 |
| Cd84 | ENSMUSG00000038147.13 | -0.31 | 0.42 | -0.73 |
| Faah | ENSMUSG00000034171.13 | -0.30 | 0.68 | -0.99 |
| Adgre5 | ENSMUSG00000002885.14 | -0.30 | 0.48 | -0.79 |
| Tspan32 | ENSMUSG00000000244.17 | -0.30 | 0.53 | -0.83 |
| BE692007 | ENSMUSG00000099757.1 | -0.29 | 1.10 | -1.39 |
| Kdm7a | ENSMUSG00000042599.8 | -0.29 | 0.52 | -0.81 |

| | | | | |
|---|---|---|---|---|
| Mta3 | ENSMUSG00000055817.17 | 0.29 | -0.33 | 0.61 |
| Pmm2 | ENSMUSG00000022711.15 | -0.28 | 0.35 | -0.63 |
| Birc6 | ENSMUSG00000024073.14 | -0.27 | 0.37 | -0.64 |
| Babam2 | ENSMUSG00000052139.18 | 0.26 | -0.39 | 0.65 |
| Ash1l | ENSMUSG00000028053.13 | -0.26 | 0.43 | -0.69 |
| Plec | ENSMUSG00000022565.15 | -0.25 | 0.46 | -0.71 |
| Stk38 | ENSMUSG00000024006.16 | -0.24 | 0.68 | -0.92 |
| 4932438A13Rik | ENSMUSG00000037270.18 | -0.22 | 0.50 | -0.72 |
| Rasa3 | ENSMUSG00000031453.16 | 0.22 | 0.86 | -0.65 |
| Snd1 | ENSMUSG00000001424.14 | 0.21 | -0.61 | 0.82 |
| Elovl6 | ENSMUSG00000041220.10 | 0.21 | -0.46 | 0.67 |
| Aff3 | ENSMUSG00000037138.17 | 0.17 | 0.87 | -0.70 |
| Itga4 | ENSMUSG00000027009.18 | 0.17 | 0.96 | -0.78 |
| Gm45552 | ENSMUSG00000110279.1 | -0.17 | 2.16 | -2.34 |
| Gpcpd1 | ENSMUSG00000027346.15 | -0.17 | 0.56 | -0.73 |
| Smg1 | ENSMUSG00000030655.15 | -0.16 | 0.44 | -0.60 |
| Bcl2 | ENSMUSG00000057329.7 | 0.16 | -0.93 | 1.09 |
| Ppp3cc | ENSMUSG00000022092.11 | 0.16 | -0.46 | 0.61 |
| Mndal | ENSMUSG00000090272.8 | 0.15 | 1.57 | -1.42 |
| Mthfd1l | ENSMUSG00000040675.17 | -0.15 | -0.98 | 0.83 |
| Scml4 | ENSMUSG00000044770.14 | 0.14 | 0.80 | -0.66 |
| Madd | ENSMUSG00000040687.16 | -0.14 | 0.46 | -0.60 |

| | | | | |
|---|---|---|---|---|
| Kmt2c | ENSMUSG00000038056.15 | -0.12 | 0.74 | -0.86 |
| R3hdm1 | ENSMUSG00000056211.13 | -0.12 | -0.78 | 0.66 |
| Myo1f | ENSMUSG00000024300.16 | 0.12 | -0.50 | 0.62 |
| Ikbke | ENSMUSG00000042349.13 | -0.11 | 0.65 | -0.77 |
| H2afy | ENSMUSG00000015937.15 | 0.11 | -0.51 | 0.62 |
| Asap1 | ENSMUSG00000022377.16 | -0.11 | 0.61 | -0.72 |
| Gm30373 | ENSMUSG00000111758.1 | -0.10 | 0.62 | -0.72 |
| Slc38a9 | ENSMUSG00000047789.5 | -0.10 | 0.51 | -0.61 |
| Airn | ENSMUSG00000078247.3 | 0.09 | 0.90 | -0.81 |
| Ncln | ENSMUSG00000020238.14 | 0.09 | 0.72 | -0.64 |
| Kmt2a | ENSMUSG00000002028.13 | 0.09 | 0.73 | -0.65 |
| Slc2a3 | ENSMUSG00000003153.10 | 0.06 | -0.65 | 0.72 |
| Tmem71 | ENSMUSG00000036944.6 | -0.06 | 0.61 | -0.67 |
| Gm26740 | ENSMUSG00000097705.1 | -0.05 | 1.13 | -1.18 |
| Il6ra | ENSMUSG00000027947.11 | 0.03 | 0.91 | -0.88 |
| H2-T24 | ENSMUSG00000053835.17 | -0.03 | 1.19 | -1.22 |
| Gm17173 | ENSMUSG00000090709.2 | 0.00 | 1.36 | -1.36 |