Genomics investigation into the regulatory network of TRIB1

Sumeet R. Deshmukh

Registration no: 180269277

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Biosciences

Department of Molecular Biology and Biotechnology

December 2021

Thesis format

1. The first chapter consists of a general introduction and background provided for each result section of the thesis.

2. Chapter 2,3, and 4 is structured as an introduction, aim, materials and methods, results, and discussion.

3. Final chapter 5 provides the conclusion of all chapters and summarizes each chapter's findings

Acknowledgment

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Ian Sudebry and Prof. Endre Kiss-toth, for their continued support of my PhD study and research, for their patience, motivation, enthusiasm, and immense knowledge. Dr. Ian Sudbery's guidance helped me in all the time research timings and thesis writing. I could not imagine having a better supervisor for my PhD journey. Besides my supervisors, I would like to thank Dr. Balint Balazs, who taught me how to think as a bioinformatician and was my first mentor in my PhD journey. Finally, Graham Hughes best manager TRAIN could ever get. A special acknowledgment Dr. Cristina Alexandru, I have learned a lot from her, especially from our science discussion. A sincere thanksto Dr. Chiara Niespolo for teaching me the lab work, always being patient with me, and incredible collaborative work. Also, I would like to acknowledge Dr. Swapna Satnam for collaborating even on the last days of her PhD. I would also like to thank the entire sudlab team for always being a slack distance away to solve any error to discuss any issues. Also, I would like to thank the entire TRAIN consortium (PI's and ESR's) for their continued support. Last but not least, I would like to thank my partner Dr. Neha Kulkarni, my family and especially my grandmother; without them, I would not have come so far.

Contents

List of figures

List of tables

List of abbreviation

- BMDMs → Bone marrow derived macrophages
- CDS → Coding sequence
- CPM → Counts per million
- DEG → Differential expressed genes
- eQTLs → Expression quantitative trait loci
- FC → Fold change
- FDR → False discovery rate
- GO → Gene ontology
- BMDM → Bone marrow derived macrophages
- INDELs → Insertion-deletion polymorphisms
- KEGG → Kyoto Encyclopedia of Genes and Genomes
- MDMs → Monocyte derived macrophages
- PBMCs → Peripheral blood mononuclear cells
- PCa → Prostate cancer
- PCA → Principal component analysis
- QC → Quality control
- RISC → RNA induced silencing complex
- SNPs → Single nucleotide polymorphisms
- TAMs → Tumour associated macrophages
- TRAIN → Tribbles research and innovation network
- TRIB → Tribbles
- UTR → Untranslated region
- WGCNA → Weighted gene co-expression network analysis
- QuASAR → Quantitative allele-specific analysis of reads
- MBASED → Meta-analysis Based Allele-Specific Expression Detection
- ASE → Allele specific expression
- AI → Allelic imbalance
- CNV -> Copy number of variations
- EMT -> Epithelial–Mesenchymal Transition
- APA -> Alternative polyadenylation
- IER -> Immediate early response gene
- Shc -> scrambled control
- DE -> Differential expressed
- LD -> Linkage disequilibrium
- OE -> Over expressed

Abstract

Advances in TRIB1 protein research reveal its potential as a biomarker for various disease diagnoses, including cancer and atherosclerosis. TRIB1 is also known as highly unstable transcript with a half-life of less than one hour, making its functional regulatory network study difficult and therefore, remains unexplored. The first part of thesis focused on understanding the regulatory network of TRIB1. In 28 different cancer datasets coexpression analysis, revealed approximately 65% of genes coexpressed with TRIB1 belong to the immediate-early response (IER) gene family. EGR1 and FOS known IER genes were present in the coexpression module in 18/28 datasests. Furthermore in-vitro analysis suggested a relationship between EGR1, FOS and TRIB1. RNA-seq analysis from early response gene stimulated TRIB1 OE prostate cancer and control (DU145) at different time-points showed effects on many of the coexpressed IER genes, Differentially expressed genes were involved in cell signaling, cell proliferation and apoptosis. Further suggested, TRIB1 could be responsible for activating genes involved in several different cellular pathways, particularly the IER pathway andTRIB1 as a member of the early response gene family,

Further study focused on understanding the post-transcriptional regulation of TRIB1 by identifying variants in the 3' UTR and their effect on miRNA binding sites The results suggests that variants in 3' UTR of TRIB1 are responsible in creating new miRNA binding sites, but they were not linked to the allele-specific expression, as would be expected if they were regulatory. The potential reasons being expression pattern of TRIB1: TRIB1 expression follows an IER pattern and degrades after one hour of cell-stress condition. In addition, novel binding sites for miRNAs generated in the TRIB1 UTR as variants are not mostly expressed in unstimulated and M1-like macrophages (LPS+ifg treated). We do however identify multiple other cases of gene variants in the 3' UTR displaying allelic imbalance in their expression.

Further study focused on understanding the role of miRNAs in macrophage polarization, nine hub miRNAs found to play an important role. For further validation, the expression of miR-125a-3p and miR-186-5p, with positive 13 control miR-155-5p were stimulated in unstimulated macrophages. RNA-seq analysis from miRNAs experiments reveals that target genes of miR-186-5p were more downregulated than the non-target ones unlike miR-125a-3p. The alternative polyadenylation analysis shows the unchanged length of 3' UTR of target genes of miR-125a-3p between unstimulated and M1-like macrophages; stating no evidence of alternative polyadenylation phenomenon for miR-125a-3p target genes and miR-125a3p has less impact on degrading its target genes.

Together these results shows that TRIB1 could a part of IER response gene network or TRIB1 is regulated by IER genes which in turn regulates the genes involved in cell-signaling, cell polarization

pathways. Next, variants in 3’ UTR of TRIB1 are not linked to ASE, however, they are responsible in altering miRNA binding sites. That could be the reason of changes in expression level of TRIB1 in different cell-type. However, this needs to be investigated further. Furthermore, not all differentially expressed miRNAs between unstimulated and M1-like macrophages are important in macrophage polarization but the subset of those miRNAs i.e., nine (hub) miRNA plays an important role in macrophage polarization.

Chapter 1

Introduction

The central dogma of biology has long been commonly known as DNA → RNA → Protein, a two-steps process involing transcription and translation. Transcription is the conversion of genetic information stored as DNA into RNA, with the maturation of messenger RNA (mRNA) transcripts being the most definitory stage of transcription. Translation is the production of the final and functional protein product from RNA, during which the ribosome scans the mRNA and uses transfer RNAs (tRNAs) to generate long peptide chains (Atkinson and Halfon 2014). (**Figure 1.1**) Moreover, the process of passing information from genes to proteins is responsible for an individual's phenotype (Alberts et al. 2017), and the complexity of gene regulation increases from single cellular to multicellular organisms (Vinogradov and Anatskaya 2007).

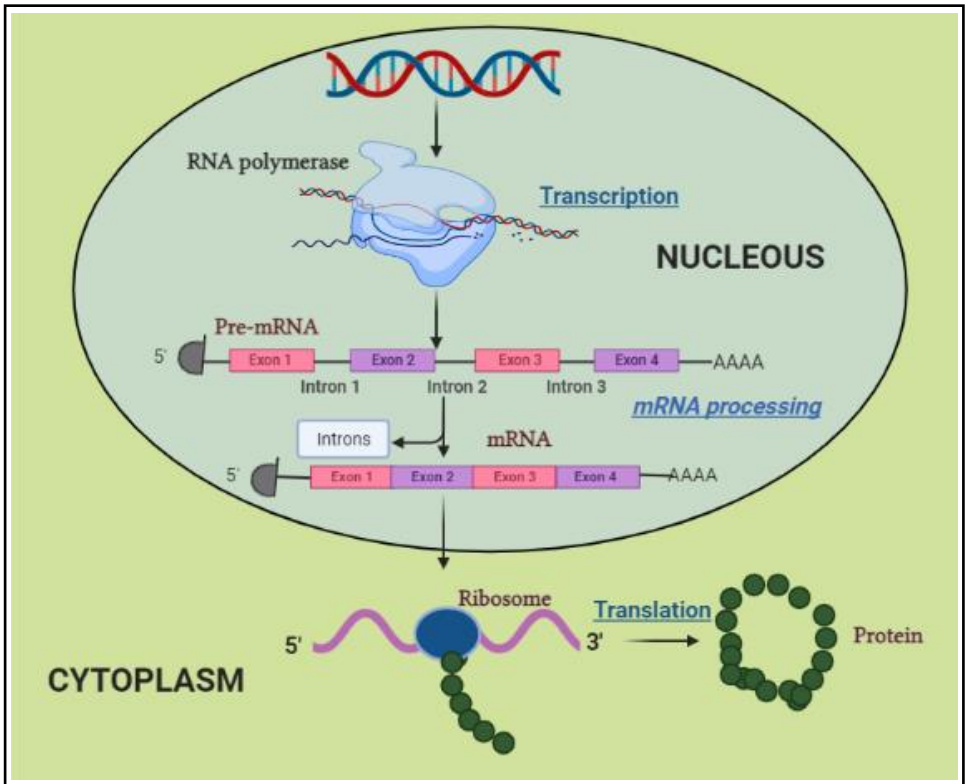


Figure1. 1- Gene expression. DNA molecule is processed by RNA-polymerase II to generate pre-mRNA (Transcription). The pre-mRNA further undergoes various steps of adding poly A tail,5' capping and finally splicing to remove introns. mRNA molecule is exported to cytoplasm from nucleus. mRNA molecule is read by ribosome to produce polypeptide chain known as protein (Translation) (Created with BioRender.com).

Regulation of gene expression has been found to occur during both transcription and translation, with mRNA transcripts containing binding sites for many regulatory elements, such as microRNAs (miRNAs) and RNA binding proteins (RBPs), which can regulate transcripts' translation, localization and stability, alike. What is more, changes in DNA sequences enhance genetic variation, can generate single nucleotide polymorphisms (SNPs) and can lead to the destruction or creation of miRNA binding sites. These can ultimately result in differences in gene expression, which has been associated with numerous diseases, including immunity-related conditions and cancer.

Thus, both miRNA and variant analyses are important for the study of the genetic architecture of complex traits. In addition, co-expression network analysis identifies interactions between genes at both genetic and protein levels (described in more detail in Section 1.6), and it can enable the understanding of specific genetic pathways directly involved with particular phenotypes and/or conditions. What is more, the use of all these analyses for the investigation of individual genes could help improve our knowledge on poorly-studied transcripts and/or genes. For example, Trib1 is a pseudo-kinase protein that has been found to be involved in macrophage polarization and differentiation, but very little is yet known about it's cellular demand, regulation and expression.

The aims of this PhD project are **1)** to investigate the genomic characteristics and regulation network of Trib1; **2)** to investigate the effects of SNPs in 3'UTR of TRIB1 on miRNA binding sites and their correlation with potential allelic imbalance in TRIB1, as well as at whole-genome scale; and **3)** to study the role of miRNAs and their expression in macrophage differentiation.

1.1 Transcription

Transcription is the first step of gene regulation and although it has been found to be more complex in eukaryotes compared to prokaryotes, the process is divided into three well-defined stages: initiation, elongation, and termination. The initiation process begins with the binding of transcription factors (TFs) to the promoter region, which has a consensus sequence that occurs across different genes, which is recognized by TFs. Before initiation can take place, a pre-initiation complex (PIC) must form on the promoter and this is composed of the RNA Polymerase II and the well-known general TFs: TF-IIA, TF-IIB, TF-IID. (**Figure 1.1**) (Cooper and Hausman 2007)

After the PIC is formed, the RNA-polymerase II is activated through phosphorylation to start the elongation step, after which other TFs such as DSIF, TFEb, and TF-IIS guide the RNA polymerase, thus starting the synthesis of the pre-mRNA from 5' to 3' direction. This is completed by adding ribonucleoside monophosphate residues to the free hydroxyl group at the 3' end of the elongated RNA chain. After elongation, the CPSF (Cleavage and polyadenylation specificity factor) binds the pre-mRNA at the poly-A (AAUAAA) region and recruits the CstF (cleavage and stimulation factor), which binds the U/GU rich region downstream of poly-A signals. CPSF and CstF interact with each other and

lead to the cleavage of the fully-matured mRNA. Further on, the XRN-2nuclear enzyme degrades the nascent mRNA to stop Pol II from synthesizing mRNA further, and thus, transcription is terminated. (**Figure 1.1**)) (Cooper and Hausman 2007).

1.1.1 RNA processing

When transcribed, an intermediary RNA known as pre-messenger RNA (pre-mRNA) is formed, which undergoes a series of processing steps, including splicing, 5' capping, and 3' polyadenylation. Splicing is mediated by the RNA-protein complex known as the spliceosome, which recognizes exons and introns' boundaries. Introns are further removed by endonucleolytic activity, and exons are joined together to form the mature mRNA. Studies have shown that transcription does not generate only one mRNA for each gene, as there are, for example, more than 75,000 mRNA encoded by 20,000 genes in the mouse genome annotation (GRCm38.p4). The process of producing multiple mRNA isoforms for the same gene is controlled by alternative splicing (AS) events, resulting in a different number of exons being included in the final mRNAs (Xu et al. 2014).

Furthermore, mRNA processing involves 5' capping, during which the terminal phosphate of the nucleotide from the transcript is removed, and guanosine triphosphate (GTP) is added at the 5' end, consisting of 7-methyl guanosine or 7 mG, a structure called 5' cap. Next, adenine residues (poly-A) are added to the 3' end of the transcript. 5' capping and poly-A tail facilitate the transfer of mRNA to the cytoplasm from the nucleus, and are also responsible for mRNA stability, as well as assisting in mRNA recognition by the translation machinery. Similarly, mRNA stability, transport and expression are regulated through miRNAs and RBPs, which can bind transcripts at different efficiencies and lead to variations in gene expression. (Wise and Lou 2021)

1.2 MicroRNAs (miRNA)

miRNAs are short, genome-derived non-coding RNAs with a role in regulating gene expression in different conditions, such as stress, infection, development etc. More than 1800 miRNA sequences have been identified in humans since the first discovery of lin-4 in *Caenorhabditis elegans* in 1993 (Hicks and Middleton 2016). Most of these regulatory elements bind to miRNA binding sites in the 3' untranslated region (3'UTR), but miRNA target binding sites have also been identified in the 5'UTRs and coding regions, although they are less effective than in 3'UTRs (Fang and Rajewsky 2011). The following sections will present an overview on the canonical pathway of miRNA biogenesis and a description of miRNA mechanism of action, particularly in regulating gene expression.

1.2.1 miRNA biogenesis

The biogenesis of a miRNA starts with the transcription of a primary miRNA (pri-miRNA), which is transcribed by RNA Pol II in the nucleus and then processed by the Drosha-DGCR8 complex that cleaves the pri-mRNA (of 100-120 nt), releases its hairpin structures, and converts it into a pre-

miRNA (70-80 nt long). The pre-miRNA is transported to the cytoplasm by Exporting-5 enzymes, where it is further processed by the Dicer-TRBP complex to double-stranded miRNA aka miRNA:miRNA star duplex (15-25 nt long). The miRNA duplex further pairs with RISC (RNA interference silencing complex), which contains argonaute protein, RNA-activator, RNA binding protein (TRBP) and DICER. Argonaute protein has two domains PIWI, which can break the RNA strand, and PAZ – the RNA-binding domain, which binds to the 3' end of the guide strand of the miRNA duplex. The PIWI domain cleaves the other strand known as the passenger strand, further degrading it during the RNA-induced silencing complex process. (O'Brien et al. 2018) (**Figure 1.2)**

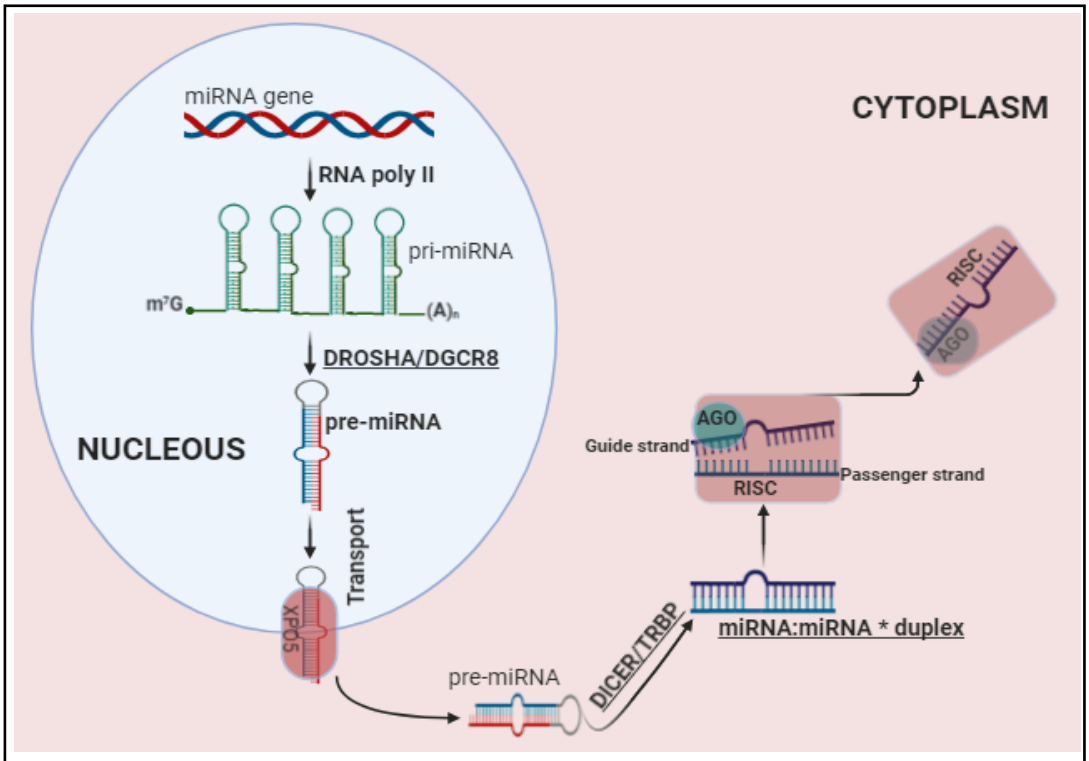


Figure1. 2- miRNA biogenesis pathway. miRNA is transcribed by Pol II into pri-miRNA and it is further processed by DROSHA/DGCR8 complex to generate pre-miRNA. Pre-miRNA is exported from nucleus to cytoplasm by the export 5 enzyme (XP05 and RanGTP). Further, DICER in complex with TRBP crops the hair loop structure of miRNAs and miRNA:miRNA* duplex is generated (~20-25 nt). Duplex miRNA is loaded in RISC complex with argonute protein. RISC cleaves the passenger strand, which is further degraded by RISC complex. The guide strand pairs with the target mRNA to induce translation repression, deadenylation and decapping of target genes. (Created with BioRender.com).

1.2.2 Gene regulation by miRNAs

Target recognition by miRNAs has been shown to be primarily focused on the pairing with the "seed" region of a binding site within the 3'UTR of a gene, which can vary in length, between 2–7 nucleotides and is usually binding the 5' end of the miRNA on the guide strand, ultimately inducing translation repression, mRNA deadenylation, and decapping (Huntzinger and Izaurralde 2011). Once target sites are identified, RISC recruits GW182, which in turn recruits other proteins, such as poly(A)-

deadenylase complexes PAN2-PAN3 and CCR4-NOT (Christie et al. 2013; Jonas and Izaurralde 2015). GW182 further interacts with PABP – a poly-A binding protein present on the poly-A tail of the target genes (O'Brien *et al.* 2018), while PAN2-PAN3 initiates the poly(A) deadenylation of the target genes until PABP can no longer bind. Subsequently, remaining adenine residues in the "terminal deadenylation" are removed by CCR4-NOT (Jonas and Izaurralde 2015), which triggers rapid mRNA degradation. The decapping of target genes is processed by the decapping protein 2 (DCP2) and associated proteins. It is then followed by the degradation of a target gene from 5' to 3' by the enzyme exoribonuclease 1(O'Brien et al. 2018) (**Figure 1.3**).

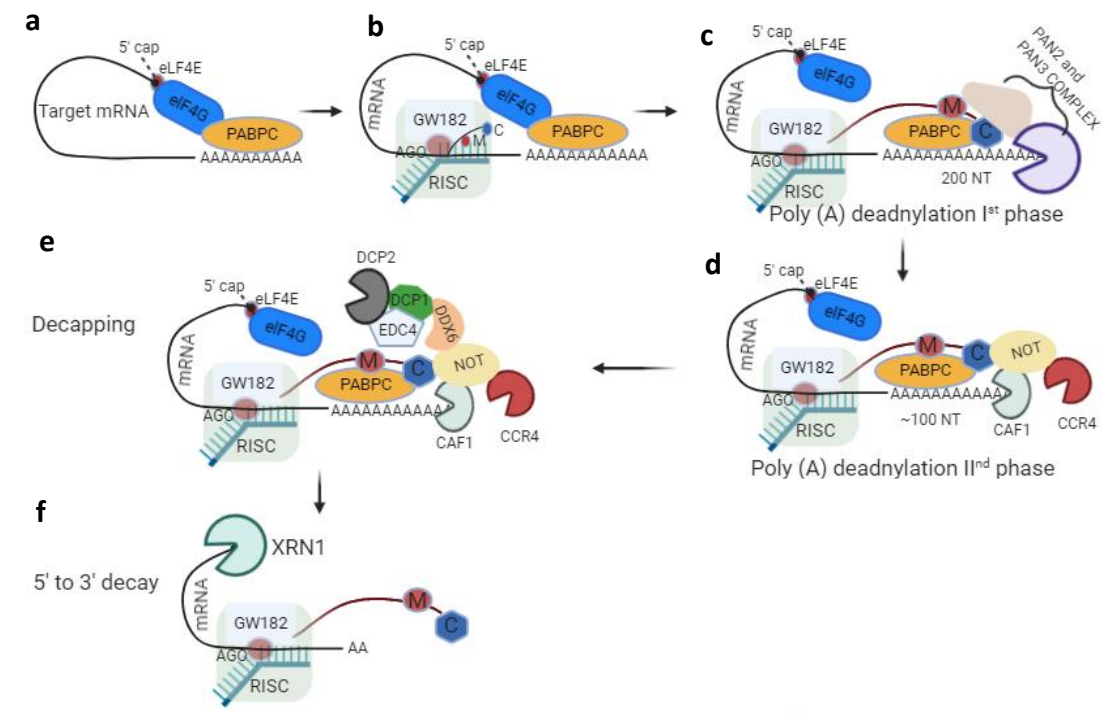


Figure1. 3- Gene regulation by miRNAs. a) target mRNA in close loop conformation formed by the interaction of cytoplasmic poly(A) with eukaryotic translation initiation factor 4F, which is further bound to eukaryotic cap binding protein (eIF4E). **b)** The guide strand of miRNA partially binds the 3' UTR of mRNA, where RISC factor recruits GW182 that interacts with AGO. **c)** RRM motif of GW182 protein then interacts with PABMC and PAN2-AN3 complex, starting deadenylation, **d)** which is continued by CCR4-NOT complex. **e)** This is followed by the decapping of target genes via decapping protein 2 (DCP2) and **f)** target mRNA degradation by XRN1 protein (Created with BioRender.com)

1.3 Bioinformatics tools for identification of miRNA targets

Although the regulation and impact of miRNAs on their targets have been well documented, their exact targeting mechanism has not yet been fully elucidated. For example, how RISC elements are recruited and how they interact to silence a target gene remains unknown. One way of studying the role of miRNAs in gene regulation is by studying the expression of both miRNAs and of their target genes. In this sense, recent advancements in next-generation sequencing technologies have proven

useful in capturing expression profiles with greater sensitivity, even for the lowest expressed genes and/or miRNAs. However, one of the most challenging and important aspect remains the identification of the bona fide mRNA targets of miRNAs, which needs to be carefully considered.

In the last ten years, numerous efforts have been made to improve miRNAs' target identification, but only a few of these have been experimentally validated. One reason for this drawback is the imperfect sequence complementarity between miRNAs and their targets, leading to incorrect identification of hundreds of genes bound by each miRNA, most of which have been later found to be false positives (not true targets) (Martin et al. 2014). However, available prediction algorithms have been developed to address this issue by applying stringent thresholds, such as using conserved miRNAs across the species, the thermodynamics stability of mRNA-miRNA interaction (Grimson et al. 2013) and seed pairing stability (Garcia et al. 2010), most importantly miRNA binding position in the 3' UTR of target genes (Grimson et al. 2013).

a) miRanda algorithm

miRanda was one of the first algorithms applied for the identification of miRNA targets in *D. melanogaster* and it was later used on the *Homo sapiens* genome (Enright et al. 2003). miRanda is based on local alignment, an adaptation of the Smith-Waterman algorithm to match the miRNA sequence at the 3'UTR of the target. It also uses an alignment scoring matrix for assigning scores and penalties for each base complementarity. miRanda also implements the Vienna package for calculating the thermodynamics folding energy of the interaction between miRNAs and the UTR (Lorenz et al. 2011). The overall false-positive rate for this algorithm was estimated at 24%, and it proved efficient in identifying 9 out of 10 currently characterized target genes (Min and Yoon 2010).

b) TargetScan

TargeScan was developed by Bartel and Burge labs at MIT in 2003, being the first algorithm to predict miRNA targets in vertebrates (Lewis et al. 2003). It requires a perfect seed complementarity between at least 6-mers of the seed sequence and 3'UTR of the target gene and its output is based on the predicted efficiency of targeting (context + scores) and the probability of conserved targeting (PCT). Thus, target conservation is first determined, followed by specific k-mer (6-mer,8-mer, 7-mer-1A) analysis. However, a 3' UTR can have multiple target sites; hence, the probability of conserved targeting (PCT) is aggregated (Riffo-Campos, Riquelme, and Brebi-Mieville 2016) and the probability of a given target being effectively bound is shown in a context+ score format. Context+ score is also based on local AU content (i.e. the AU content of a transcript), 30 nt upstream and downstream of the predicted site, position contribution (i.e. the distance to the nearest end of an annotated target UTR), the 3' pairing with the seed sequence and the miRNA-target complementarity outside the seed

sequence (Grimson et al. 2007). The false-positive rate for TargetScan was estimated between 21% and 31% (Agarwal et al. 2015).

c) PicTar

PicTar was developed by Rajewsky's group in 2005 (Krek et al. 2005) and compared to miRanda and TargetScan, it focused mainly on 7nt long sequences as seed sequences, starting from position 1 or 2 of the 5' end of a miRNA. It also allows one mismatch, one deletion, or one insertion for base pairing between miRNAs and 3' UTR of the target gene. The algorithm starts by retaining all targets conserved across species, followed by a calculation of the free energy for the entire miRNA:mRNA duplex, after which the target sites are filtered based on a free energy threshold provided. Lastly, the PicTar score – a maximum likelihood score is computed from the hidden Markov Model (HMM). This algorithm has not been updated since its development, but it has proven efficient in identifying true miRNA targets, with a false positive rate of about 30% (Min and Yoon 2010).

In addition to above several other miRNA prediction tools or webservers are available to identify miRNA target genes such as miRanalyzer (Hackenberg et al. 2009), Sylamer (van Dongen, Abreu-Goodger, and Enright 2008), Co-expression Meta-analysis of miRNA Targets (CoMeTa) (Gennarino et al. 2012), miRTarCLIP (Wang et al. 2013),TarPmiR (Ding, Li, and Hu 2016).

For the purpose of this project we implemented widely used algorthims such as miRanda and TargetScan algorithm.

1.4 Regulatory genetic variation

The differences in DNA sequences within the genomes of different individuals of the same species are known as genetic variations (Ku et al. 2010). These, together with variations in gene expression can be associated with disease onset, progress and severity, hence understanding the correlation between gene expression and gene variants can improve our knowledge on the genetic architecture of complex traits. Generally, it has been observed that most genetic variations have little-to-no impact on any disease or trait, with many neutral mutation changes in DNA sequences not affecting an organism 's chance of survival or fitness.

A silent mutation is a change in DNA sequence without a subsequent change in the amino acid, but some of the resulting transcript variants may lead to disease, such as in the cases of melanoma (Gartner et al. 2013) , schizophrenia (Duan et al. 2003). In contrast, certain variants may provide a positive advantage in changing environments: for example, CCR5 gene – the primary co-receptor used by HIV envelope protein gp120 sequentially with CD4, located on chromosome 3 carry the Δ32 mutation (Galvani and Slatkin 2003), which was found to be responsible for generating a premature stop codon that produces a non-functional protein (Ni, Wang, and Wang 2018). Individuals homozygous for this mutation (with a Δ32/ Δ32 allele) in CCR5 show no expression of this protein on

T-cell receptors. This leads to a loss in the target/binding site of HIV glycoprotein 41 on the T-cell receptor, making these individuals resistant to HIV1 infection (Lopalco 2010).

1.4.1 Genetic variants in microRNA (miRNA) binding sites

Due to the specificity of miRNA binding on mRNAs, introducing a single SNP in the 3'UTR of its target may lead to functional changes. For example, SNPs in the seed sequence can either completely remove existing miRNA binding sites, or can create novel ones, ultimately affecting gene expression. More than that, SNPs in the seed sequence may also affect the efficiency of miRNA and mRNA pairing, which in turn may increase or decrease post-transcriptional regulation of target genes. Given the number of SNPs in *Homo sapiens,* it comes as no surprise that SNPs in the seed sequence of miRNA binding sites have shown an effect on protein expression and its association to disease.

1.4.2 SNPs in the seed sequence and the creation of miRNA binding sites

Numerous SNPs in seed sequences of miRNA binding sites have been found to be responsible for creating new binding sites. For example, Clop *et al.* found that a variant in the 3' UTR of GDF8 in sheep presented a point position from G to A and thus created a new binding site for mir-1 and mir-206. These, in turn, reduced the GDF8 gene expression, leading to Muscle Hypertrophy (Clop et al. 2006). Similarly, studies have shown an A>C SNP (rs4245739) in the Mdm4 3' UTR, where C in the minor allele created a new binding sites for miR-191 and/or miR-887-3p, which ultimately led to a decreased protein expression (Che, Shao, and Li 2014; Lin et al. 2014; Yang et al. 2017). This was particularly important, as overexpression of Mdm4 was previously associated with cancer development, by negatively regulating the p53 tumor suppressor protein (Wade, Wang, and Wahl 2010).

Another example of variation in tumour suppressor proteins is the identification of two isoforms of TP63: TAp63 and ΔNp63, which result from alternative splicing and the use of different promoters(Westfall and Pietenpol 2004). TP63: TAp63 was shown to be a tumour suppressor, while ΔNp63 an oncogene (Park et al. 2013). A study from Wang *et al.* identified a C>T SNP (rs35592567) in the 3'UTR of ΔNp63 that created a binding site for miR-140-5p, which decreased protein levels and thus lowered the risk of bladder cancer (Wang et al. 2016).

APOC3 gene encodes for C3 protein that is involved in triglyceride metabolism along with lipoprotein lipase (LPL) and apolipoprotein C2 (APOC2) (Hu et al. 2016), and the overexpression of C3 was found to increase the risk of coronary heart disease (CHD). However, individuals with a G>T SNP (rs4225) in the 3' UTR of the APOC3 gene showed a decrease in the risk of CHD and further investigation revealed that a T on the minor allele created a miRNA binding site for miR-4271, leading to decreased translation of APOC3 (Caussy et al. 2014; Hu et al. 2016).

1.4.3 SNPs in the seed sequence and the loss of existing miRNA binding sites

Numerous SNPs in seed sequences of miRNA binding sites have also been shown destroy already-existing binding sites. For example, a A T>C SNP (rs10719) in the Drosha 3' UTR, located close to the binding site of miR-27b, affected the pairing efficiency of miR-27b and the Drosha transcript, which increased its expression (Yuan et al. 2013). The Drosha enzyme plays an essential role in miRNA biogenesis: it cleaves the pri-miRNA and liberates the hairpin structure, converting it into pre-miRNA and its overexpression was found to increase the risk of bladder cancer (Zhang et al. 2015).

Another well-known example of SNPs degrading miRNA binding sites was found in Cluster of Differentiation 86 (CD86 or B7-2), which is present on the surface of antigen-presenting cells and acts as a key mediator of T-cell identity (Geng et al. 2014). It plays an essential role in autoimmunity, transplantation, tumour immunity, and it was found to lead to dysfunctional immune cell populations when expressed at high levels. Such variations in CD86 were are also proven responsible for T-cell deactivation, unresponsiveness to tumour cells, and increasing the production of the pro-inflammatory cytokine interleukin 4 (IL4), which was associated with high risk of colorectal cancer (Kwasniak et al. 2019). A G>C SNP (rs17281995) in the 3' UTR of CD86 was predicted to disrupt five miRNA binding sites, out of which miR-582 was found to be less efficient in binding to the minor allele, leading to an increase of CD86 expression levels (Geng et al. 2014).

Similarly, the HIF-1α protein, encoded by the HIF1A gene, is a transcriptional factor activated in response to oxygen deprivation and influences cell metabolism and cell survival, and it plays an important role in ischemic diseases and tumour angiogenesis. The SNP (rs2057482) T>C in 3'UTR of HIF1A was found to reduce the binding efficiency of miR-199a, thus increasing its gene expression. More than that, pancreatic ductal adenocarcinoma patients with a CC (alternate) genotype in HIF1A were observed to have larger tumour sizes, followed by those with heterozygous CT and homozygous TT genotypes (Guo et al. 2015).

Another representative example is the T>C SNP (rs4937333) detected in the 3'UTR of EST1, which impaired the binding of miR-5003-3p, enhances the differentiation of B cells into plasma cells (Zhang et al. 2019). This was particularly important, as aberrant innate immune responses dysregulate cytokine production, and aberrant activation of B cell differentiation into plasma cells plays an important in the pathogenesis of Systemic lupus erythematosus (SLE, or lupus) (Choi, Kim, and Craft 2012). A Genome-wide association study (GWAS) has identified genetic alterations in many genes associated with SLE (Mohan and Putterman 2015), with EST1 being one of the most preponderant one, which was found to be responsible for reducing peripheral blood mononuclear cells (PBMCs) in SLE patients (Yang et al. 2010).

1.5 Bioinformatics tools for detection of regulatory variants

As described above, detecting SNPs and/or transcript variants with regulatory potential is useful in accurately associating specific genes and their expression with particular disease or phenotypes. Different strategies have been developed in this sense, with identification of expression quantitative trait loci aka eQTL being a widely used method for studying the impact of variation within a genomic region on transcript regulation (**Figure 1.4**). An eQTL identifies the loci that affect the expression level of mRNA, by studing the genomic variants and transcriptomic data from the same

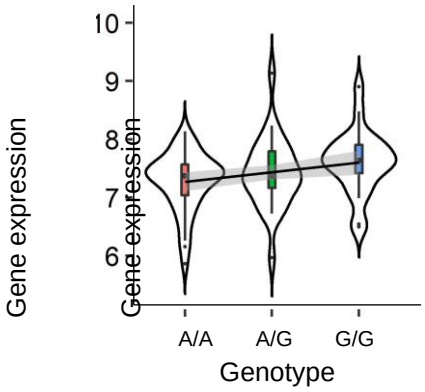


Figure1. 4- Example of eQTL. Violin plot showing a linear relationship between dependent variable i.e., gene expression on y-axis and independent variable i.e. genotype on x-axis, where A is a reference allele and G is an alternate allele. The horizontal line represents a linear relationship between three genotypes.

individual (Jung et al. 2020). One of the first studies performed by Goring *et al.* in 2007, using peripheral blood lymphocytes, identified an eQTL located in the promoter region of the vanin1 (VNN1) gene, which was found to influence high-density lipoprotein (HDL) cholesterol concentrations. This study demonstrates the use of eQTL for associating a potential gene with an affected human traits (Göring et al. 2007).

eQTLs are divided into two types: cis-eQTL or local eQTLs, in which a genetic variant is present within one megabase (Mb) from the transcription start site (TSS) of affected genes; and trans-eQTL, where a genetic variant is located on a different chromosome or away from the affected genes (Qin, Liu, and Xie 2021).

Cis-eQTLs can regulate gene expression by affecting several steps in the gene expression process, including mRNA splicing, TF binding, DNA methylation and miRNA binding. For example, a cis-eQTL can be a variant present in the promoter region or enhancer, affecting transcription factor binding. A two base pair deletion in the promoter sequence of the ERG28 gene is such an example,

which reduced the binding affinity of two TFs – SOK2 and MOT3, ultimately decreasing the expression of EGR28 in yeast (Chang et al. 2013). Many studies have identified cis-eQTLs across the whole genome, such as the one performed by Urmo Vosa *et al.* on 3186 blood samples, which reported 88% of genes (16,989) have an associated cis-eQTL (Võsa et al. 2018). Moreover, GTEx – probably the biggest eQTL study to date, reported that 94.4% of protein-coding genes and 67.7% of lincRNA genes have at least one cis-eQTL in at least one of 49 tissues (François Aguet, Kristin G. Ardlie 2020).

Trans-eQTLs are variants located in different genes, but disturbing the expression of multiple genes across the chromosomes. What is more, many studies have shown that trans-eQTL can also be associated with nearby genes, including with cis-eQTLs. For example, Bryois *et al.* used 869 lymphoblastoid cell lines from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort and identified cis-eQTLs for TFs BATF3 and HMX2 that disturbed multiple trans-genes across different chromosomes (Bryois et al. 2014). Such eQTLs are called eQTL hotspots, and they are mostly present near major regulators or master TFs (Breitling et al. 2008).

Another example is the trans-SNP rs1215608 located within the NUAK1 gene in skin cells, shown to be associated with three different genes *MO6P*, *PPM1F*, *LECT1.* Such trans-SNPs are also known as multi-gene regulators (Grundberg et al. 2012), and unlike cis-eQTLs, these are often present across tissues, while trans-eQTLs tend to be cell-specific (Dimas et al. 2009). For example, Raj *et al.* performed eQTL profiling of $CD4^+$ T cells and monocytes derived from blood of 461 individuals with ancestry from different continents. They found 32% of cis-eQTLs are specific to monocytes, 8% to T cells and 62% were shared between two cell types. In contrast, 482 trans-eQTLs were associated with 55 genes specific to monocytes, 31 genes specific to T-cells and only four genes were shared between the cell types (Raj et al. 2014).

Although eQTL analysis can be used to successfully identify variants affecting gene expression, it is very expensive, and it requires a large number of samples, ranging from hundreds to thousands, as well as matched whole-genome and transcriptome sequencing data. It may be challenging to achieve such a large data set, specifically for tissues obtained via surgical procedures (e.g. atherosclerosis plaques). It is also worth mentioning that for measuring the genotype and gene expression to perform eQTL mapping requires excellent expertise and many *in vitro* experiments, which are also time-consuming.

1.5.1 Allele-specific expression (ASE)

Another approach to identifying the impact of genetic variation on gene expression is allele-specific expression (ASE). Diploid organisms carry two alleles for each gene, inherited from each parent and the expression of these alleles is responsible for an individual's phenotype (Buckland 2004). The ratio of expression between the two alleles is usually balanced at 50:50 for an expressed

gene, but in few instances it was found that the expression of one allele is relatively higher than that of the other, which led to allelic imbalance, also known as differential allelic expression (Maia et al. 2009) or allele-specific expression (ASE) (Liu, Dong, and Li 2018a) (**Figure 1.5**).

Unlike traditional eQTL mapping, ASE analysis has greater statistical power for detecting cis' genetic elements, even with small sample numbers (Harvey et al. 2015). For example, Liu *et al.* aimed to identify the allele-specific expression responsible for tumour initiation and progression, and they used normal and cancer tissue samples from 59 individuals with Colorectal Cancer (CRC). They identified 50 tumor-specific ASEs that contributed to the somatic events in the genes' regulatory regions and in significantly enriched cancer-driving genes. This study demonstrates that allele-specific expression events can be captured with relatively few samples (Liu, Dong, and Li 2018b).1

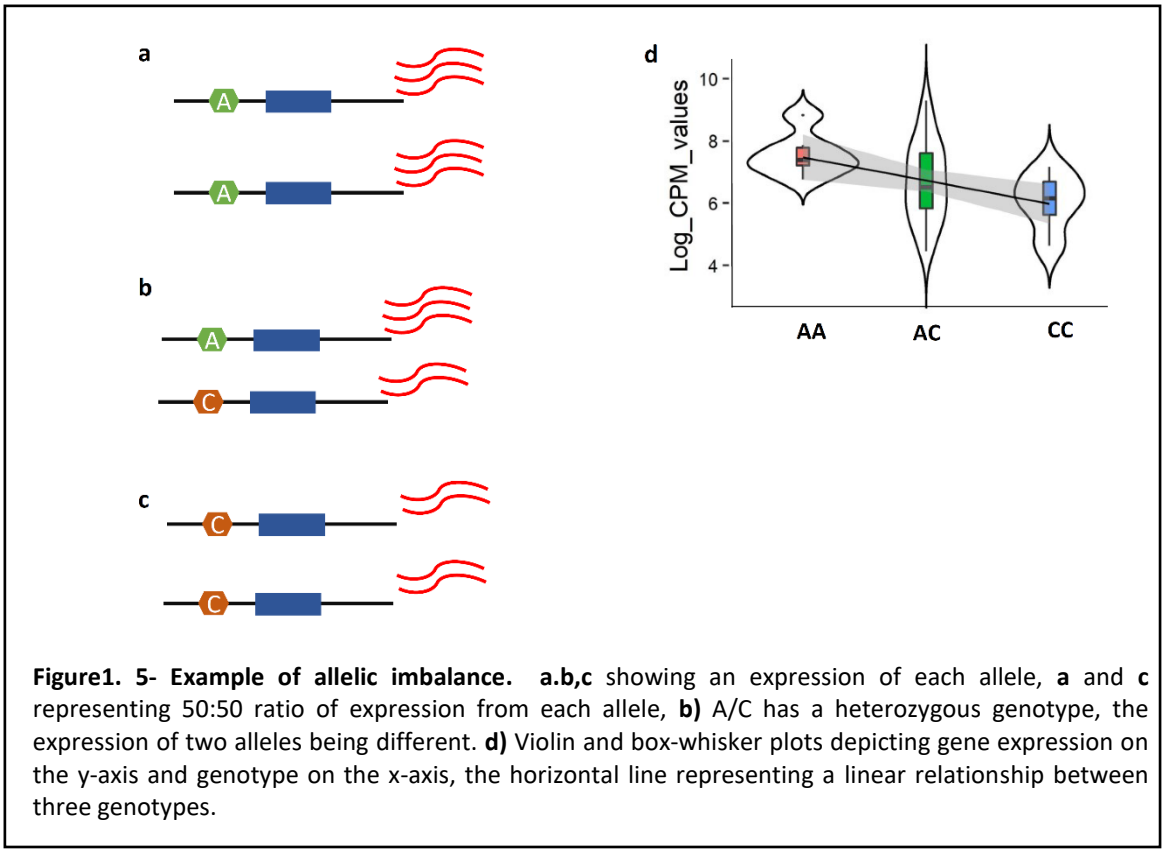


Figure1. 5- Example of allelic imbalance. **a.b,c** showing an expression of each allele, **a** and **c** representing 50:50 ratio of expression from each allele, **b)** A/C has a heterozygous genotype, the expression of two alleles being different. **d)** Violin and box-whisker plots depicting gene expression on the y-axis and genotype on the x-axis, the horizontal line representing a linear relationship between three genotypes.

In the past decade, there has been a consistent progress in the study of allele-specific expression using high throughput sequencing such as RNA-seq (Mortazavi et al. 2008), which enables the measurement of individual gene expression in various tissues, with greater sensitivity and specificity than the micro-array technique (Rao et al. 2019). RNA-seq also allows capturing ASE instances by quantifying the difference between one parental allele's expression over the other (Castel et al. 2015). Asymmetric expression of two alleles of the same gene is the potential evidence of cis-regulatory expression, or cis-eQTL and signature of an allele-specific expression. However, in order to identify allele-specific expression, RNA-seq data needs to have an adequate read depth, sufficient heterozygous sites, and to be aligned correctly to the respective genome.

1.5.1 Tools for allele-specific expression analysis

A number of pipelines and software, designed using different platforms and approaches have been developed for the detection, evaluation and analysis of allele-specific expression, some of the most widely used being described below.

a) Alleleseq

Alleleseq was developed by the Gerstein Lab at Yale University for the study of ASE. This software requires both the individual's RNA-seq or ChIP-seq data and phased variation information (box 3), including SNPs and CNVs of each parent obtained from trio-sequencing. Using this method, a personalized diploid genome is constructed based on phase information used to map the individual's sample. The next step is the number of reads mapping to each maternal variant being compared to the paternal variant *via* a binomial test. Multiple testing correction is then applied to eliminate false positives. However, the statistical test does not take into account the replicate information (Rozowsky et al. 2011).

b) Allim

Allelic imbalance meter (Allim) requires both RNA-seq data from offspring, as well as genomic information, SNP information, DNA-seq or RNA-seq of the parents. Based on the parent genomic information, it generates a diploid genome that is then used for mapping the individual's RNA-seq sample. ASE is counted based on the reads that are uniquely mapped to either of the parent genomes. It is observed that some genes have more reads mapped to one of the parents, which leads to mapping bias. Hence, to remove the bias, a similar number of reads are randomly generated for both parents, and mapping bias is further estimated for each gene and corrected for allelic expression. The statistical test (no biological replicates: G-test and replicates: ANOVA test) is then used to remove false-positive results (Pandey et al. 2013).

c) MBASED

Meta-analysis based allele-specific expression detection (MBASED) uses only RNA-seq data to identify allele-specific gene expression, and it is predominantly implemented using R programming language. To identify ASE at gene level, it combines allelic frequency information of heterozygous SNPs within a gene across multiple individuals. If prior haplotype information is not provided, then MBASED uses the pseudo-phasing approach, in which the larger read count of the allele at each SNP is assigned as a major haplotype. However, when haplotype phasing information is present, it increases the power of ASE detection. MBASED combines the allelic frequency across individual heterozygous SNPs within a gene and detects deviation within the sample. The advantages of using this tool is its fast speed compared to other tools, the requirement for less memory and it capacity to handle SNPs across a whole transcriptome with multiple samples.(Mayba et al. 2014)

d) ASEP

Allele-specific expression analysis in a population (ASEP) tool identifies gene-based allele-specific expression under one condition and differential ASE between two conditions, using only RNA-seq data from a given population of individuals. This approach can be applied to samples with unknown haplotype information, but read depth should be high for haplotype phasing reconstruction. If read depth is low, then using known haplotype information increases ASE detection's power. ASEP uses the same approach as MBASED for constructing haplotype phasing information, but using a higher number of SNPs increases the memory requirement. (Fan et al. 2020)

e) QuASAR

The Quantitative allele-specific analysis of reads (QuASAR) tool was also developed using R language and it was initially implemented for RNA-seq data, but it can also be applied on DNA-seq, ChIP-seq, ATAC-seq. In this method, heterozygous SNPs are identified and ASE is counted accurately by considering base-calling errors and overdispersion in the ASE ratio. QuASAR is the first method in identifying genotypes and ASE from the same sequencing data (Harvey et al. 2015).

For the purpose of this current project only MBASED and QuASAR were used, which will be described in more detail in Chapter 3.

1.6 Co-expression analysis

One of the objectives in biological studies is to identify molecules present in each system and to determine their interaction at both genetic and protein levels. The function of many genes is yet to be understood, but a common assumption is that genes involved in the same cellular pathway may be co-expressed and co-regulated. One of the strategies often used to investigate such interaction networks between genes is guilt-by-association, which implies that genes associated with or interacting are more likely to share functional characteristics (Cui and Churchill 2003). However, for this technique to be successfully applied, the functions of many other proteins need to be known.

Before RNA-seq or microarray technologies, researchers used to carry multiple experiments on various cell lines to understand the regulatory network of a particular disease or an unknown gene’s function. Although this approach has successfully identified the regulatory mechanism of many diseases and functions of several novel genes, it remains expensive and time-consuming (Yu et al. 2012). Since the cost of RNA-seq has significantly decreased, it has become an increasingly popular technique for measuring gene expression across multiple conditions/phenotypes at the same time.

However, the RNA-seq technique is often used to identify differentially expressed genes between two states, e.g. diseased and control, by calculating changes in the mean level of expression

of each gene (Cui and Churchill 2003). It is assumed that genes with a shift in their mean levels are associated with a given disease phenotype, but most of the information in the gene expression dataset is ignored. Moreover, sometimes genes known to be involved in a particular disease are not statistically significantly altered between conditions (Hudson, Dalrymple, and Reverter 2012). This could be due to the disease state being caused by a change in function (e.g. *via* a mutation in either the coding or untranslated region), without an actual effect on its gene expression.

As an alternative, co-expression analysis approaches can handle many of the above limitations (Brazhnik, De La Fuente, and Mendes 2002), the two most well-known methods being clustering-based and network-based co-expression analysis.

1.6.1 Clustering-based co-expression analysis

In the late '90s microarray techniques became popular for studying the expression pattern of thousands of genes in different conditions simultaneously (Do and Choi 2008), several different clustering methods being developed to represent groups of genes with similar expression patterns (Belacel, Wang, and Cuperlovic-Culf 2006). It was assumed that genes clustering together based on their expression patterns would often share biological functions.

Wen and colleagues performed one of the first co-expression analyses, testing this hypothesis by generating mRNA expression data from 112 rat cervical spinal cord tissue, using reverse transcription-coupled PCR. They then clustered genes based on their expression patterns, identifying 5 main clusters, with genes in each cluster share indeed similar biological functions (Wen et al. 1998). In another independent study of the same year, M. Eisen and colleagues analysed 8,600 transcripts from human fibroblasts cells, with and without serum starvation, using microarray technology. They, too, observed that genes were grouped by their expression patterns in 5 different clusters, and each cluster contained genes with similar biological functions (Eisen et al. 1998).

Furthermore, Spellman and colleagues combined time-series expression data from yeast, also obtained *via* DNA-microarray technology, with the genes' promoter sequences. They also observed genes clustering together were coregulated, and they identified many common regulatory factors related to cell cycle. (Spellman et al. 1998). In addition, H. Lee and colleagues analysed 60 different human datasets containing 3924 microarrays, on which they performed a correlation analysis between each pair of genes, ultimately identifying a high-confident network of 8805 genes, which were then confirmed to correlate with functional relatedness, through further functional analysis (Lee et al. 2004).

As co-expression and co-regulation analysis became more common in high-throughput sequence analyses, numerous new clustering methods were proposed, including self-organized map clustering (SOM) (Kohonen 1982), hierarchical clustering (Eisen et al. 1998), k-means clustering

(MacQueen 1967) and bi-clustering (Pontes, Giráldez, and Aguilar-Ruiz 2015). It is worth mentioning that the clustering-based approach identifies or formulates clusters based on expression patterns, without considering biological relevance. This increases the chances of forming a cluster with a set of co-expressed genes and not co-regulated, e.g. the different conditions could trigger the expression of genes that can show a similar expression to different groups of genes with diverse biological functions (Pirim et al. 2012).

1.6.2 Network-based co-expression analysis

In biology, networks can be classified based on the compounds and interactions represented, e.g. in co-expression networks, the compounds are genes, and the interactions show shared expression patterns, whereas in protein-protein interaction networks, the compounds are proteins and interactions represent physical interactions. In the co-expression network, genes are considered nodes, and the edges are determined by their correlation (Serin et al. 2016). Network-based co-expression analysis is usually performed in three steps, detailed below, in **Box 1**.

Box 1: Network analysis or co-expression network analysis procedure is as follows:

- **First step:** Similarity score (explained in box 2) is calculated by comparing gene expression patterns in pairs;
- **Second step:** The result list of gene pairs is filtered using the similarity score threshold;
- **Third step:** The remaining pairs form nodes and corresponding edges to construct a network. Genes with similar similarity scores form a sub-network or modules, or cluster. According to the guilt by association principle, genes sharing a biological function tend to have similar expression patterns (Wolfe, Kohane, and Butte 2005). Known genes can predict the function of an unknown gene in the same module (Rhee and Mutwil 2014).

The values in a co-expression-based co-regulation network are usually between +1 and -1. An unsigned network is based on the absolute value of correlation, so the pairs of strong-negatively correlated genes are considered to be co-expressed or highly connected. This approach treats the positive and negatively correlated genes equally. In contrast, for the signed network only correlation values between 0 and 1 are used. Hence, the pairs of strong-negatively correlated genes are considered not connected or not co-expressed, as their connection strength is very close to or zero. Biologically-relevant modules are often identified *via* the signed method (Mason et al. 2009).

A weighted network assumes all genes are connected in the network, and the connection between genes is assigned a continuous value between 0 and 1. The higher the value, the stronger the strength of co-regulation between the genes. In contrast, an unweighted network assumes that genes are either connected or unconnected, as the interaction between them is binary (0 or 1) (Zhang and Horvath 2005).

For the purpose of this current project, we focused on signed weighted co-expression network, by using Weighted gene co-expression network analysis (WGCNA) tool, also known one of the most widely used computational tools for constructing and analyzing co-expression networks. As mentioned above, it uses soft thresholding to build a network, and a topological overlap matrix aka TOM (**Box 2**) is applied with hierarchical clustering on the correlated gene expression values, to create co-expression modules (Langfelder and Horvath 2008). These modules are further refined by cutting the branches given at certain heights, and co-expressed genes and the hub genes can be inferred without prior knowledge of their structure or function. Hub genes are a set of genes with high correlation to the largest number of the other members of the module. WGCNA was the first method to be applied on RNA-seq datasets and has since been considered the most efficient tool for identifying biologically relevant associations between phenotypes and modules (Kogelman and Kadarmideen 2014).

1.6.3 Weighted gene co-expression network in transcriptomic data

Network-based co-expression analysis has been implemented in many gene expression studies (Wu et al. 2019), such as the one performed by M.S. Cao and colleagues, who analysed mRNA expression data of normal, gastric carcinogenesis and adenocarcinoma samples from the GEO database (GSE24375), and generated a regulatory network for each condition. Thus, they identified 16 hub genes and hypothesized the role of GATA6, ESRRG and their signaling pathways in gastric cancer development (Cao et al. 2015).

Box 2

Similarity score:

A similarity score is calculated on the log2 transformed value of gene expression. There are number of methods available for calculating similarity scores. E.g., Simple Pearson or Spearman rank correlation are widely used methods. Although Pearson assumes data is normally disturbed and has linear relationship. Spearman rank correlation is more robust but less powerful (Ballouz, Verleyen, and Gillis 2015). Another method often used is mutual information (MI), which can be applied to non-linear datasets (Meyer, Lafitte, and Bontempi 2008). However, a comparison between MI and Pearson correlation shows no difference in results (Song, Langfelder, and Horvath 2012).

Threshold values:

A user-specified threshold similarity value is used in practice, which is considered connected and the remaining unconnected, i.e., hard thresholding (Zhang and Horvath 2005). E.g., In the signed network, values less than 0.5 could demonstrate a negative correlation. In contrast, values greater than 0.5 would be associated with positively correlated genes (Mason et al. 2009). While (Langfelder and Horvath 2008) propose to use soft thresholding implemented in the WGCNA model. The idea behind soft thresholding is raising correlation to a power (user-defined) will reduce the correlation's noise.

Topological overlap matrix (TOM):

TOM measure is a matrix product of adjacency matrix with itself. In addition, TOM normalize the results of adjacency matrix between 0 to 1. Adjacency matrix considers pairs of genes only connected to each other, while TOM pairs of genes are connected to other set of genes in the network (Yip and Horvath 2007).

Similarly, in a study by Yi Lio and colleagues WGCNA was applied to generate the co-expression network on mRNA expression data for lung adenocarcinoma (LUAD) samples obtained from The Cancer Genome Atlas (TCGA) database. Overall, they identified five hub genes CHEK1, RAD51, KIF18B, KIFC1, FEN1, and RAD54L that strongly influence LUAD stem cell maintenance, and were involved in cell cycle and DNA replication pathways. These hub genes were further validated by analysing six more expression data sets of LUAD patients and healthy individuals (controls) obtained from the GEO database. They again observed that all six hub genes were highly expressed in LUAD tissues and tumor recurrence patients, but none of these were validated experimentally (Liao et al. 2020).

Components linked to diabetes-associated cardiovascular disease were also investigated through network analysis by W. Liang and colleagues, who also used corresponding tissue data sets from the GEO database (GSE13760). Statistically significant genes were used to construct a co-expression network using WGCNA, and they identified a total of 15 hub genes that were present in all modules. The hub genes were further screened against the GWAS database (cardiovascular disease catalog), and three genes, HLA-DRB1, LRP1 and MMP2 were identified, which were found to play an essential role in antigen-presenting and phagosome pathways, intracellular signalling and lipoprotein metabolism associated with cardiovascular disease, inflammation, tissue remodeling (Liang et al. 2020).

1.7 Macrophage differentiation

Humans are surrounded by toxic and allergenic pathogens such as bacteria, fungi and viruses, which can attack and replicate inside the human body, leading to tissue damange and affecting beneficial, commensal microbes. The host immune system is, however, well equipred for eliminating foreign biological threats *via* two major interacting responses: the innate (non-specific response) and adaptive (specific response) immunity (Chaplin 2010).

Innate immunity is the first line of host defense against harmful pathogens by initiating a protective inflammatory response, which aims to destroy any infectious microorganisms and remove dead or dying cells and damaged extracellular matrix materials. It is also involved in repairing damaged tissues and restoring them to a healthy condition (Matzinger 2007). On average, the innate immunity is activated between 4-96 hours from infection or tissue stress, and it continues untill the infection is cleared or an adaptive response is activated (Kaur and Secord 2019).

The mononuclear phagocyte system (MPS) plays an important role in the inflammatory response, removing infected dead cell debris, and repairing damaged tissues. The MPS consists of different cell types, such as monocytes, tissue residential macrophages and dendritic cells (DCs) (Hume, Irvine, and Pridans 2019). Growth factors and different types of cytokines, such as interleukins regulate the function of MPS cells, help maintain homeostasis (tissue morphology and tissue function), drive proliferation and are involved in differentiation (Hume and MacDonald 2012).

1.7.1 Macrophage differentiation

Monocytes are a group of cells circulating in blood that can be mobilized to the spleen or lungs on-demand, and they are approximately 10% of the total nucleated cell. (Ginhoux and Jung 2014). They originate in the bone marrow from hematopoietic stem cells (HSCs) and can circulate in the blood stream for 1-2 days if no danger is identified (Italiani and Boraschi 2014). Upon infection with foreign pathogens, monocytes can differentiation into dendritic cells or un-stimulated (M0) macrophages. The latter can be subdivided into three categories, based on the expression of CD14

and CD16 surface markers (Ziegler-Heitbrock et al. 2010): a) 90% of total cell count are made up of monocytes with high expression of CD14 and no expression of CD16 ($CD14^{++}CD16^{-}$), which is known as classical macrophage, while the remaining 10% represent a minor population of monocytes, further divided into b) intermediate, characterized by high CD14 and low CD16 levels ($CD14^{++}CD16^{+}$) and c) monocytes with relatively low CD14 and high CD16 expression levels ($CD14^{+}CD16^{++}$), also termed as a non-classical subset (Ziegler-Heitbrock et al. 2010). The classical and intermediate subset of human monocytes exhibit a pro-inflammatory phenotype, also known as M1 macrophages, while non-classical monocytes have an anti-inflammatory phenotype / an M2 polarisation state.

Initially, macrophages were hypothesised to differentiate from monocytes circulating in blood in response to injury (Tauber 2017), but it was later found that monocyte-derived macrophages are originated from bone marrow, while lineage-tracing studies have shown that tissue-resident macrophages are derived from embryonic yolk sac progenitor and Langerhans cells (Hoeffel et al. 2012). These tissue-resident macrophages are named based on their location, as described in **Table 1.1**, and despite their different origins and locations, they all have similar functions, such as in homeostatic maintenance and tissue repair after injury.

Table 1 1 Different types of macrophages

Resident macrophages	**Location**
Kupffer cells	Liver
Sinus histiocytes	Lymph nodes
Microglia	Central nervous sysem
Hofbaur cells	Placenta
Osteoclasts	Bone
Epithelioid cells	Granulomas

Monocytes differentiate into pro-inflammatory macrophages upon detection of infectious microorganisms by pattern recognition receptors (PRRs) such as Toll-like receports (TLRs) (Nau et al. 2002). Thus, they enter an M1 polarisation state and release specific cytokines, such as IL12, TNF, IL-6, and chemokines like CCL2, CXCL10, CXCL11, TNF- α, IL-1α. Also, they produce chemicals such as reactive oxygen species (ROS) and nitrogen radicals (caused by upregulation of inducible NO synthase iNOS), which enhances their ability to kill pathogens. M1 macrophage genes are regulated by transcription factors such as NF-kB, STAT1, STAT5, IRF3 and IRF5 (Murray 2017). M1 macrophages are also important antigen-presenting cells (APC), and after killing the infectious pathogens, they display

major histocompatibility complex (MHC) class I and II antigens, further activating the acquired immune system.

Monocytes can also differentiate into M2 or anti-inflammatory macrophages, which are induced by cytokines like IL4 and IL13 and are specific cells of adaptive immunity (Viola et al. 2019). They also have a high expression of CD206, IL-R (decoy receptor), and IL-1R antagonist, which stimulate pro-fibrotic factors, such as TGF-β and IGF-1 and in turn reduce inflammation and promote tissue repair (Mantovani et al. 2013). Additional genes expressed in M2 macrophages are STAT6, GATA3, CD163, FIZZI1, MMPs, PPARY and ARG1 (de Groot and Pienta 2018), whose increased expression leads to the production of polyamines and promotes tissue remodeling and healing. Moreover, M2 macrophages are known to induce angiogenesis (growth of blood vessels from the existing vasculature) and lymphangiogenesis (growth of lymphatic vessels from the existing lymphatic vessels), by stimulating the vascular endothelial growth factor A an IL-8 (Corliss et al. 2016). Given their various properties and characteristics, M2 macrophages can be divided into different groups, as described in **Table 1.2**.

Table 1 2- Types of resident macrophages and their locations

Macrophage	Stimulation	Cytokines	Markers	Function
M1	LPS+IFN-ϒ	IL6,IL12,IL23	CD80,CD86,MHC-II	Pro-inflammatory activity
M2a	IL4, IL13	TGF-β, IGF-1	CD206,CD36	Anti-inflammatory activity,
M2b	TLR	IL10,IL6,TNF-α	CD86, MHII	Tumour progression, T-helper cell response.
M2c	IL10	IL10,TNF-β	CD163,TLR1	Removal of dead cells, tissue remodeling
M2d	TL3	IL10, VEGF	CD206,CD204	Angiogenesis, lymphangiogenesis, tissue remodeling

1.8 Tribbles

Pseudo-kinases are proteins containing kinase-like domains, but lacking one or more conserved motifs required for efficient ATP binding and/or catalysis in their canonical counterparts (Taylor and Kornev 2010). They are known to be involved in cell signaling by mediating protein-protein interactions (Boudeau et al. 2006), and have been associated with various human diseases, including auto-immune diseases (Ribeiro et al. 2019), metabolic and neurological disorders and different types of cancer (Bailey et al. 2015).

Tribbles are a family of serine/threonine pseudo-kinases consisting of 3 members – TRIB1, TRIB2, and TRIB3, which are involved in numerous regulatory pathways in eukaryotic cells. TRIB2 was the first member to be characterized and was initially known as C5FW. It was first identified in dogs and linked to an upregulated thyroid response to mitogens (Wilkin et al. 1997). TRIB3 was first described by Mayumi-Matsuda *et al.* in 1999, and its increased levels were linked to neuronal cell death induced by nerve growth factor depletion. TRIB1 (C8FW) was subsequently described as a close homolog of TRIB2 by Wilkin's group. The sequence similarity of Tribbles protein is shown in **Table 1.3**.

Table 1 3- Sequence similarity between Tribbles proteins

Pairwise protein sequence similarity	% similarity
TRIB1 and TRIB2	71.3%
TRIB1 and TRIB3	53.3%
TRIB2 and TRIB3	53.7%

1.8.1 Structural features and function of Tribbles protein

Evolving from a common ancestor, all members of the tribbles family contain three common domains: N-terminal PEST domain, central kinase-like domain, and C-terminal protein-binding domain. The protein's central domain resembles a serine/threonine kinase domain, but it lacks three major key motifs of kinase catalytic activity: 1) a VAIK motif, known to interact with α and β phosphate of ATP, 2) a central HRD domain, which contains the catalytic aspartic acid residue known for base acceptor required for protein transfer, and 3) a DFG domain, which binds MG2+ to coordinate β and ϒ phosphate of ATP in binding (Hegedus, Czibula, and Kiss-Toth 2006). Therefore, Tribbles have been classified as pseudokinases, with TRIB2 and TRIB3 having low ATP affinity and phosphotransferase capacity, while TRIB1 lacks the ability to bind ATP and exert phosphotransferase activity.

The C-terminal domain has two conserved motifs: 1) an HPW[F/L] motif, which contains binding sites for MEK1 and MAPKK that, in turn, mediate the binding of kinase 3 (MLK3) which is involved in c-Jun N-terminal kinases (JNKs)-signaling pathways, and the p53 mitogen-activated protein kinase (MAPK) signaling pathway (Humphrey et al. 2010; Yokoyama et al. 2010); and 2) the DQXVP[D/E] motif, which allows binding of E3 ubiquitin ligase COP1 (**Figure 1.6**) (Keeshan *et al.* 2010). TRIB1 and TRIB2 recruit COP1 to degrade CEBPα and β, responsible for developing acute myelogenous leukemia in murine models (Keeshan et al. 2010), and for inhibiting the AKT protein responsible for suppressing adipocyte differentiation (Naiki et al. 2007). TRIB3 also recruits COP1, resulting in the degrading of acetyl coenzyme carboxylase (ACC) in adipose tissues (Qi et al. 2006).

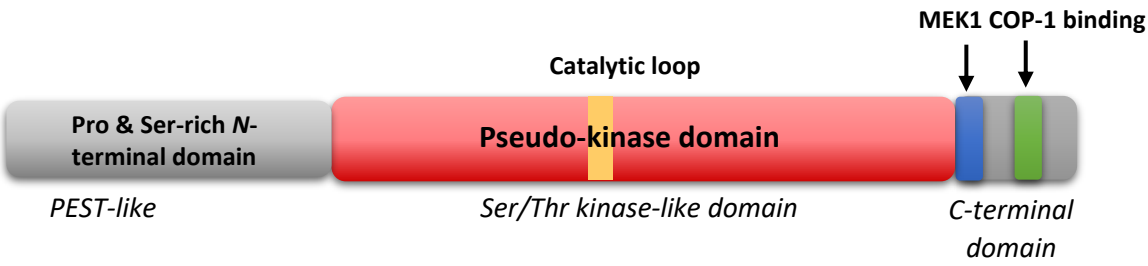


Figure1. 6- Basic structure of Tribbles. All three members of this protein family contain variable N-terminal PEST domain, psedokinase doman lacking metal binding motif (DGF motif) and C-termain domain containing two key regulatory elements: HPW[F/L] motif interacting with MAPK/MEK family proteins, and DQXVP[D/E] conserved motif, binding E3 ubiquitin ligase COP1.

All three homologus Tribbles members Trib1, Trib2 and Trib3 play important roles in various cellular processes, including inflammation, cellular stress, apoptosis and tumorigenesis. They have also been found to be involved in the differentiation of macrophages (Satoh et al. 2013), muscle cells (Kato and Du 2007), adipocytes (Sathyanarayana et al. 2008) and osteoblast (Chan et al. 2007).

1.8.2 TRIB1

The TRIB1 mRNA transcript is considered an unstable transcript, with a half-life of less than 1 hour, whereas the average median half-life of most stable human transcripts is at least 7 hours (Sharova et al. 2009). However transcript half-life depends on its physiological conditions (Newbury 2006), e.g most of the housekeeping genes have long mRNA transcript half-lives, while proteins required for a limited period of time or involved in different stages of the cell cycle, cell differentiation or growth have shorter half-lifes (Sharova et al. 2009). The expression of TRIB1 is highly variable and cell-type specific, which suggests that TRIB1 might be subjected to post-transcriptional regulation (Soubeyrand, Martinuk, Lau, et al. 2016). What is more, the length of 3'UTR of TRIB1 is approximately three times longer than that of its coding region (**Figure 1.7**), which means there is a larger genetic

space in which mutations can occur and can, for example, lead to alteration of miRNA binding sites and affect transcript's stability.

Figure1. 7- Genomic view TRIB1

TRIB1 is primarily expressed in antigen-presenting and endothelial cells (Ashton-Chess et al. 2008), while in HeLa cells it was identified as a mediator of gene expression and a regulator of the Toll/IL-1 receptor in innate immunity (Kiss-Toth et al. 2004). Moreover, TRIB1 plays a vital role in the polarization of vascular smooth muscle cells *via* inhibition of the MAPK–JNK pathway (Hye et al. 2007). In addition, the expression of TRIB1 was found to be increased in patients with atherosclerotic arteries compared to control individuals (Johnston et al. 2019). Another study, performed by E. Dugast *et al.*, suggests that in T-regulatory cells TRIB1 shows binding affinity to FOXP3 (Dugast et al. 2013) , which is a TF important in regulating T-cells (T-regs), promoting long-term survival of organ transplant patients (Li and Turka 2010), while deficiency of FOXP3 has been associated with auto-immune diseases and inflammation.

Furthermore, TRIB1 was also shown to have roles in fatty acid metabolism, as hepatic over-expression of TRIB1 in mice resulted in increased fatty acid oxidation, decreased plasma triglyceride and cholesterol, by reducing very low-density lipoprotein production (Bauer et al. 2015). Another complementary study in TRIB1 knockout mice recorded an increased expression of triglyceride and cholesterol (Soubeyrand, Martinuk, Naing, et al. 2016). Other studies have emphasised additional roles of Trib1, such as the 8q24 GWAS locus containing TRIB1, which has been associated with coronary artery disease (Jadhav and Bauer 2019), and TRIB1 expression was found to be upregulated in LPS-induced white adipose tissues (Ostertag et al. 2010). What is more, heterozygous knockout TRIB1 mice have impaired cytokine (IL-6 and IL-1β) production in white adipose tissues and are protected from weight gain and adiposity when fed with a high-fat diet (Ostertag et al. 2010).

i) Role of TRIB1 in macrophages

As discussed above, TRIB1 shows binding with FOXP3 TF, which is a master regulator of T cells' activation, and TRIB1 deficiency impairs cytokine gene expression in white adipocytes. Moreover, Arndt and colleagues reported that TRIB1 play an essential role in the differentiation of tissue-resident M2-like macrophages, as they aimed to understand the impact of TRIB1 deficiency on macrophage function and polarisation. They also observed that bone marrow-derived macrophages from TRIB1-deficient mice showed a reduced expression of both M1 (IL6, ILb, and Nos2) and M2 markers (Cd206, Fizz1, and Arg1) upon LPS/IFNY and IL-4 stimulation, respectively. Ultimately, they

found a reduction in pro-inflammatory cytokine, nitric oxide, and reactive oxygen species levels (Arndt et al. 2018). In addition, TRIB1 was also found to be involved in macrophage migration through interaction with C/EBPβ and TNF-α, while in TRIB1 knockout samples, TNF-α, IFN-γ and/or TLR2 ligands were recorded at high levels (Liu et al. 2013).

Another study performed on mice lacking TRIB1 in hematopoietic cells reported reduced adipose tissue mass and increased lipolysis, even when fed a normal diet. However, this condition was rescued after injecting the mice with M2-like macrophages. In contrast, when the mice were fed a high-fat diet, they developed hypertriglyceridemia and insulin resistance (Satoh et al. 2013). Despite the numerous studies focused on TRIB1, the mechanisms leading to impaired macrophage polarisation or the exact genetic and/or protein-protein interactions involving TRIB1 have not yet been elucidated.

ii) Role of TRIB1 in cancer

In addition to its role in macrophage differentiation and polarisation, TRIB1 has also been identified as an oncogene in acute myeloid leukemia (AML). As discussed above, TRIB1 recruits E3 ligase COP1 to degrade target C/EBPα and it interacts with MEK1, enhancing ERK phosphorylation to promote cell proliferation and inhibition of apotosis, processes that have been associated with pathogenesis of adult acute myeloid leukemia (AML) (Yokoyama et al. 2010) . It was also observed that overexpression of TRIB1 is responsible for increasing Hoxa9-induced leukemia, by modifying the activity of a super-enhancer of the oncogene Erg, which is achieved through the degradation of C/EBPα p42 by TRIB1. BRD1 inhibitor JQ1 has been shown to inhibit the same super-enhancer, thus impairing the growth of TRIB1-expressing AML cells (Yoshino et al. 2020). Moreover, TRIB1 is located on chromosome 8q24, 1.5 Mb away from c-MYC, a known oncogene associated with AML, which was detected in low levels in previous studies, while TRIB1 was concomitantly overexpressed. This led to the hypothesis that TRIB1 may also have a cooperative role with c-MYC. (Storlazzi et al. 2006).

In addition, TRIB1 was found responsible for inhibiting the tumor suppressor protein p53 (Miyajima, Inoue, and Hayashi 2015), a TF with vital role in regulating genes involved in cell-cycle progression and apoptosis (Kruse and Gu 2009). What is more, a study by Ying Ye and colleagues suggested that TRIB1 might be a target for miR-23a, which is positively regulated by p53. In the absence of miR-23a, TRIB1 expression is increased, which would further inhibit p53. They also suggested that hepatocellular carcinoma (HCC) cell migration (Pérez et al. 2020) and EMT activation are promoted by ectopic expression of Trib1 *via* upregulating β-catenin, which is responsible for upregulation of c-Myc and MMP-7. They further reported that the effect of TRIB1 on β-catenin was p-53 dependent and the overal results indicated a potential role of Trib 1 in wound healing and initiation of metastasis in cancer progression (Das et al. 2019; Ye et al. 2017).

TRIB1's association with different types of cancer has been documented in numerous studies, the most prominent ones linking TRIB1 overexpression in prostate cancer (Niespolo et al. 2020; Shahrouzi et al. 2020), thyroid cancer (Puskas et al. 2005), ovarian cancer (Puiffe et al. 2007) and colorectal cancer (Liang et al. 2013). However, the expression pattern of TRIB1 in different cancer datasets (different no. of cancer and control samples – **Table 1.4**) is variable, as shown in **Figure 1.8**.

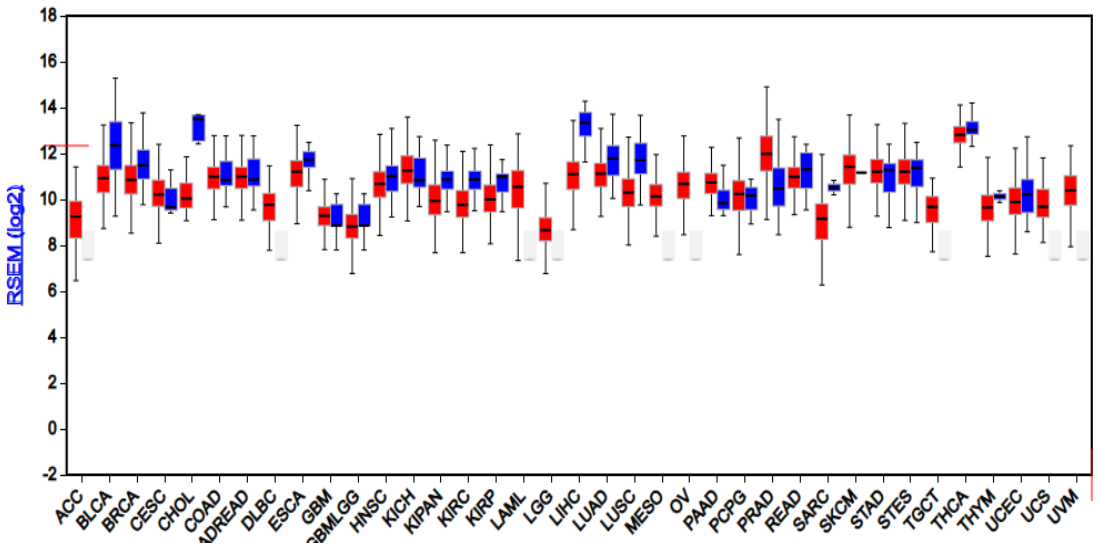


Figure1. 8- Variation of TRIB1 in different cancer datasets. Box-whisker plot showing the TRIB1 mRNA expression (RSEM) in different types of cancer (red) and their respective normal samples (blue) . The number of samples for each tissue-specific dataset are shown in **Table 1.4** (Firehose Broad GDAC (*Broad GDAC Firehose (broadinstitute.org)*

Table 1 4- Number of cancer and control **samples for each tissues-specific dataset**

Cancer type	Abbreviation	Number of cancer samples	Number of normal samples
Adrenocortical carcinoma	ACC	79	NA
Bladder urothelial carcinoma	BLCA	408	19
Breast invasive carcinoma	BRCA	1100	112
Cervical and endocervical cancers	CESC	306	3
Cholangiocarcinoma	CHOL	36	9
Colon adenocarcinoma	COAD	459	41
Colorectal adenocarcinoma	COADREAD	656	51
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	48	NA
Esophageal carcinoma	ESCA	185	11
Glioblastoma multiforme	GBM	166	5
Glioma	GBMLGG	696	5
Head and Neck squamous cell carcinoma	HNSC	522	44
Kidney Chromophobe	KICH	66	25
Pan-kidney cohort (KICH+KIRC+KIRP)	KIPAN	891	129
Kidney renal clear cell carcinoma	KIRC	534	72
Kidney renal papillary cell carcinoma	KIRP	291	31
Acute Myeloid Leukemia	LAML	173	NA
Brain Lower Grade Glioma	LGG	530	NA
Liver hepatocellular carcinoma	LIHC	373	50
Lung adenocarcinoma	LUAD	517	59
Lung squamous cell carcinoma	LUSC	501	51
Mesothelioma	MESO	87	NA
Ovarian serous cystadenocarcinoma	OV	307	NA
Pancreatic adenocarcinoma	PAAD	179	4
Pheochromocytoma and Paraganglioma	PCPG	184	3
Prostate adenocarcinoma	PRAD	498	52
Rectum adenocarcinoma	READ	167	10
Sarcoma	SARC	263	2
Skin Cutaneous Melanoma	SKCM	472	NA
Stomach adenocarcinoma	STAD	415	35
Stomach and Esophageal carcinoma	STES	600	46
Testicular Germ Cell Tumors	TGCT	156	NA
Thyroid carcinoma	THCA	509	59
Thymoma	THYM	120	NA
Uterine Corpus Endometrial Carcinoma	UCEC	546	35
Uterine Carcinosarcoma	UCS	57	NA
Uveal Melanoma	UVM	80	NA

iii) Overview of TRIB2 and TRIB3

Compared it its other members of the Tribbles family, TRIB2 has more diverse actions, being highly expressed in kidney mesenchymal cells, and considered responsible for suppressing adipocyte differentiation by inhibiting AKT and C/EBPβ (Naiki et al. 2007). In contrast, downregulation of TRIB2 was found to regulate LPS-induced IL-8 production *via* the MAPK pathway (Eder et al. 2008).

TRIB3 is the most studied isoform of the Tribbles family, being reported in high levels in liver samples, and relatively elevated levels in the small intestine, stomach, kidney, lung and white adipose tissue (Okamoto et al. 2007). The expression of TRIB3 was shown to be upregulated in response to cellular stress, including oxidative stress, hypoxia, glucose excess and essential amino acid deficiency (Ord and Ord 2017). What is more, the proximal promoter region of TRIB3 has a C/EBP–ATF composite binding sites that mediates binding of C/EBP, CHOP, C/EBPβ, and C/EBPγ, leading to TRIB3 upregulation in cellular stress response (Huggins et al. 2016). TRIB3 is also regulated by other transcriptional factors such as PPARα, FoxO1 (Koo et al. 2004; Matsumoto et al. 2006), the former being a significant regulator of hepatic fatty acid, which binds to the TRIB3 promoter region and regulates its expression in liver, while FoxO1 is identified to bind to the promoter region of TRIB3 and alter its expression in murine samples ((Ord and Ord 2017).

Box 3

Early response genes

Regulation of gene expression is an important process used by cells to control the end products of genes. Specific group genes are known to respond quickly to internal or external stimuli such as cell stress conditions like viral or bacterial infections. These group genes are known as immediate early response genes (IEGs). Hence, they are known as first responders having a peak expression within 30 minutes after external or internal stimuli. The main function of IEGs is to regulate cell growth and involve different cellular processes (Bahrami and Drabløs 2016).

Haplotype phasing

With the increase in technology vast amount of genotype data is being generated; these data identify alleles co-located on the chromosome. However, we can't observe which of the chromosomes or haplotypes belong to either of the parents. Identifying the alleles belonging to either of the parents is known as haplotype phasing (Browning and Browning 2011).

1.9 Aims and Objectives

In the following chapters, both computational analyses, using methods like the ones described above, as well as experimental validation (where applicable and available) will be presented for each of the three main sub-projects, focused on investigating: **1)** the genomic characteristics and regulation network of Trib1; **2)** the effects of SNPs in 3'UTR of TRIB1 on miRNA binding sites and their correlation with potential allelic imbalance in TRIB1, as well as at whole-genome scale; and **3)** the role of miRNAs and their expression in macrophage differentiation.

The genomic characteristics and regulation network of Trib1 chapter focuses on understanding the regulatory network of TRIB1 through coexpression analysis and hypothesizing the function of TRIB1**.** As discussed above variations in expression level of TRIB1 led it difficult to understand its function and regulatory network. From the published literature, we understand that TRIB1 is involved in different types of cancer and involved in cell signaling such as MAPK signaling pathways, this function makes TRIB1 biologically important gene. In order to understand the function and regulation of TRIB1, we have analysed 28 different cancer datasets using (WGCNA). We observed that 65% of genes co-expressed with TRIB1 are early response genes, these genes were further validated using RNA-seq dataset. From RNA-seq data we found that TRIB1 is regulated by early response genes (box 3) and it play a role in regulating cellular response in stress

The effects of SNPs in 3'UTR of TRIB1 on miRNA binding sites and their correlation with potential allelic imbalance in TRIB1, **as well as at whole-genome scale chapter** focuses on understanding the effect of mutations in 3' UTR of TRIB1 on miRNA binding sites, also we expanded this analysis to identify if these SNPs in 3' UTR of TRIB1 are linked to the allele-specific expression of TRIB1. As discussed in the introduction, the length of 3' UTR of TRIB1 is three time longer than its coding region and 3' UTRs are more prone to mutations, which could led in the alteration in miRNA binding sites or different MRE factors such as RNA-Binding proteins (RBPs), translational repressors, splicing factors, and riboswitches. This could be the reason for variations in expression level of TRIB1. In order to investigate the mutation in TRIB1 responsible in alteration in miRNA binding sites, we analysed unstimulated macrophage datasets and identified that mutations in 3' UTR (using haplotypecaller GATKS's tool) are responsible for change in expression level of TRIB1, however, they were not responsible in alteration of miRNA binding sites expressed in macrophages, but could be altering binding sites of other MRE factors. In addition to this, we also investigated if mutation in 3'UTR of TRIB1 linked to allele-specific expression, we developed allelic specific expression pipeline, implementing QuASAR for identifying SNPs linked to ASE and MBASED to allele specific expressed genes. From our results, we did not find any SNPs of TRIB1 linked to ASE.

The role of miRNAs and their expression in macrophage differentiation chapters focuses on understanding the role in miRNA in polarizing M0 to M1 macrophages. micro-RNAs (miRNAs) are small, non-coding RNA sequences binding to un-translated regions (UTRs) of protein-coding genes and are responsible for regulating transcript expression, stability and transport. In order to understand the role of miRNA in macrophage differentiations, we performed integrated miRNA-mRNA analysis between resting (M0) macrophages and M1-polarized macrophages. Differential expression analysis of miRNA and mRNA were performed independently using DESeq2 R package. We found that up-regulated target genes of down-regulated miRNAs are involved in inflammatory related pathways and downregulated target genes of upregulated miRNAs plays an important role in cell-cycle related pathways. Moreover, we identified 5 upregulated miRNA and 4-downregulated miRNAs targeting more than 500 genes and these genes are enriched in inflammatory and cell-cycle related pathways. We further validated our results, analysing RNA-seq data, obtained from downregulated hub miRNAs in M0 macrophages. In addition to this, we also investigated alternative polyadenylation sites of differentially expressed genes using DaPars algorithm and we observed that upregulated genes undergo APA process to avoid the binding sites of upregulated miRNAs.

Chapter 2

Co-expression analysis of TRIB1

2.1 Introduction

Over the past decade, the application of massively parallel sequencing at large scales has led to an increase in research and reporting of new information on cancer genomics. One such project that benefited from these technological advances has been the one studying TRIB1, as one of the members of the Tribbles pseudokinase family. As mentioned in the Introductory Chapter 1, TRIB1 was recorded as upregulated in many different cancer datasets (Niespolo et al. 2020; Puskas et al. 2005; Wang et al. 2017; Yoshino et al. 2021) and it was also associated with cell-cycle-related pathways, which control cell polarization, mobility, differentiation, and other cellular functions relevant to cancer biology (Eyers, Keeshan, and Kannan 2017). Despite the clinical significance of TRIB1, its only relationship validated experimentally, through TRIB1 knockdown in mice, is that with COP1 and CEBPα (Yoshida et al. 2013) . Until now, the genomic regulatory network and full extent of TRIB1 functions have remained unexplored. Therefore, this study focuses on understanding the genomic regulation and targets of TRIB1 through co-expression network analysis.

Continuous advancements in RNAseq technologies have improved the understanding of the complexity of the human genome and have enabled the identification of differentially expressed genes between two states/conditions (e.g. disease and control, or gene knockdown and over-gene expression). However, studies on the genome-wide immunity response in TRIB1-knockout mice, carried out by collaborators in the TRAIN consortium failed to find large transcriptional effects linked to TRIB1 in resting, steady-state cells (data not shown). Therefore, we implemented a co-expression network based approach to identify the regulatory network of TRIB1.

This method clusters genes in different modules, based on their correlation between their gene expression values across samples, and builds networks that can be used to understand the interaction between these genes. Using this strategy, possible roles of unknown genes can be inferred based on similar biological functions of the known genes present in the same module (Eisen et al. 1998). More than that, after identifying a module of interest, gene enrichment, pathway analysis and Gene Ontology (GO) analysis can help unravel the entire module's biological function. Ultimately, a set of co-expressed genes can be further validated experimentally, in order to understand and/or confirm the relationship between known and unknown genes. Furthermore, in order to determine the direction of relationship and causality between genes, which cannot be inferred from co-expression networks alone, in vitro and in vivo validation are essential (**Figure 2.1**).

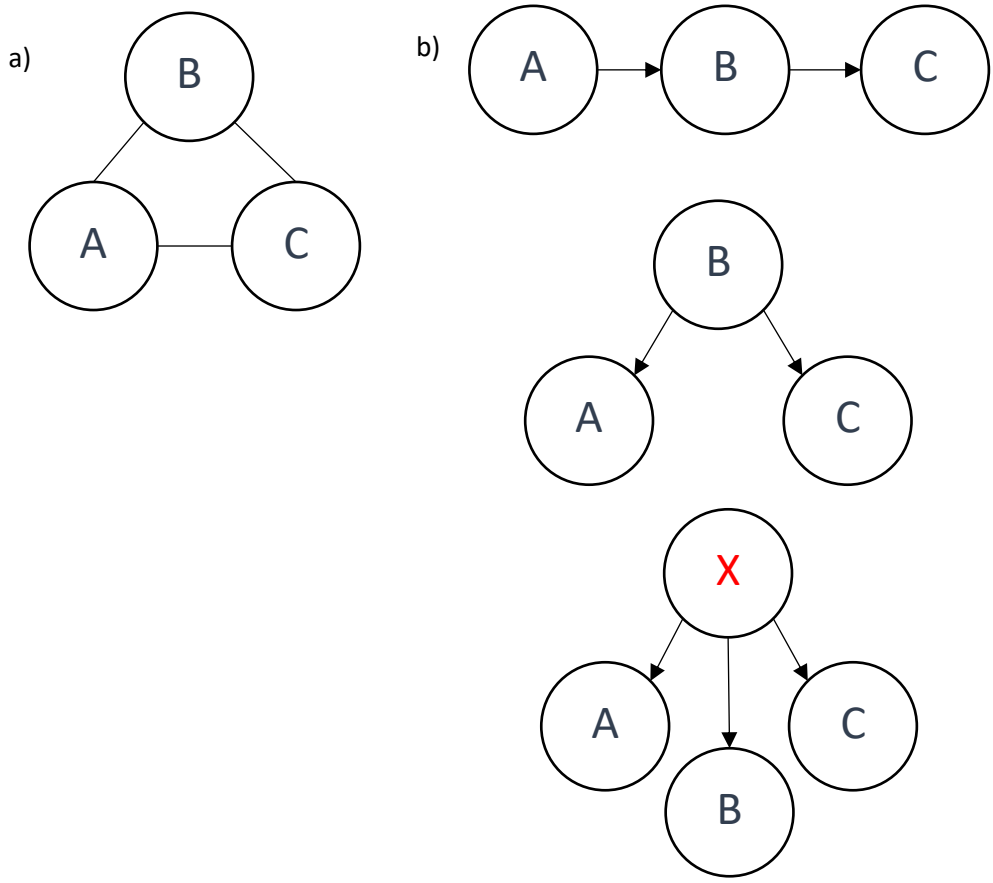


Figure 2. 1 - Differences between co-expression network and regulation network. a) Network built from co-expression analysis, b) Possible causal relationships

Aims and objectives

In this study, we harvested network modules generated by the weighted co-expression method (WGCNA) in order to identify genes co-expressed with TRIB1 in macrophages across 28 different cancer datasets. We aimed to investigate the regulatory network of TRIB1 through co-expression analysis, and to generate a hypothesis on its function. Based on the co-expression network results, we then hypothesized that TRIB1 could 1) be an early response gene, 2) regulate early response genes, or 3) be regulated by early response genes, which we further tested *in vitro*.

2.2 Materials and Methods

Declaration

Stable DU145 cell lines overexpressing TRIB1 and PC-3 cell lines with TRIB1 knockdown were generated by our collaborator, Dr. Swapna Satnam, under the supervision of Dr. Natalia Pellegata (Helmholtz Institute, Munich, Germany) and Prof. Endre Kiss-Toth (IICD, University of Sheffield, Sheffield, UK). She also performed the time-series experiments on the DU145 cells. Dr. Chiara Niespolo performed qPCR of EGR1 and FOS in TRIB1 knockdown conditions in PC-3 cells. All other experiments on HEK293T cells, computational analyses, and interpretation of results were performed by myself.

2.2.1 In vitro methods

i) Generation of stable cell lines, with TRIB1 overexpression or TRIB1 knockdown

TRIB1-overexpressing and TRIB1-knockdown cells were derived through transformations using lentivirus encoding a TRIB1 expression cassette. HEK293T, DU145 and PC-3 cells were grown overnight, in Dulbecco's Modified Eagle Media (DMEM) + 10% Foetal Bovine Serum (FBS), but without antibiotics, to a concentration of 1 X 10^6 cells/well in 6-well plates. The following day, cells were transduced with lentivirus and supplemented with 8 µg/ml Polybrene. After 6 hours, culture medium without antibiotics was added to the cells and these were incubated for another 24 hours. The following day, Puromycin-containing media was added and then subsequently changed every 2-3 days. Stable cell lines were frozen and stored for further processing.

For immediate-early response stimulation experiments, cells were serum-starved for 12 hours, in just 0.5% FBS. To stimulate early response genes, 20% FBS was added, and cells were collected at 0, 30, 60, 90, 120 and 180 mins time points.

ii) Total RNA-extraction

Total RNA from HEK293T cells was extracted using RNAeasy Mini kit (Qiagen), according to the manufacturer's instructions, and RNA quality was assessed by NanoDropTM Spectrophotometer (A260/A280 ratio of 1.8-2.0) (ThermoFisher Scientific). Total RNA from DU145 was isolated using RNA-tissue kit (Promega) and Maxwell® and RNA quality was tested using Nanodrop 2000, checking for an A260/280 ratio of 1.8-2.2.

iii) cDNA synthesis and qPCR

cDNA synthesis for HEK293T and PC-3 cell line was performed using iScript cDNA synthesis kit (Bio-Rad) according to manufacturers instructions, and for DU145 High-Capacity RNA-to-cDNA™ Kit (Thermo Fisher Scientific) was used.

Quantitative Real-time PCR (qPCR) was performed using PrecisionPLUS SYBR-Green master (Qiagen) mix (for HEK293T and PC-3 cells) or TaqMan (for DU145 cells), and EGR1, FOS, and TRIB1 primers (**Table 2.1**). All experiments were performed with technical replicates and analysed on CFX384 C1000 Touch Thermal Cycler (Biorad), values being normalized to the housekeeping gene GAPDH. For analysis, the below steps were followed:

1. The average technical replicate Ct value was calculated for each sample
2. The Ct value of GAPDH was subtracted from those of individual genes, to calculate ΔCt
3. ΔCt of control samples was further subtracted from test samples, ΔΔCt
4. $2^{-\Delta\Delta Ct}$ was calculated, by raising 2 to the power of negative ΔΔCt

Table 2. 1 - qPCR primers used

Gene	Forward primer	Reverse primer
TRIB1	AAGTTCACATTTGAACTGATGGC	AGCTGGTTTCAGGGGAAGAC
EGR1	GAAGAACTTGGACATGGCTGTTTC	CCTCCCTCTCTACTGGAGTGGAA
FOS	CCAACCTGCTGAAGGAGAAG	AGATCAAGGGAAGCCACAGA
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC

iv) siRNA transfection

2,50,000/well HEK29T and PC-3 cells were seeded in 6-well plates, 12 hours prior to transfection, and grown in standard growth medium (DMEM + 10% FBS), and cells were then serum-starved for further 12 hours by replacing 10% FBS containing media with media without FBS. Uniform distribution of cells and healthy cell density were confirmed prior to transfection for both cell lines using microscopy, and Viromer Blue (Lipocalyx) (Protocol is discussed in box 1) was used to transfect siRNAs. After 24 hours of siRNA transfection, PC-3 cells were collected, while for the H2K293T cells, 20% FBS was first added, and cells were collected after 1 hour.

v) RNAseq from time-series experiment on DU145 cells overexpressing TRIB1

DU145 cells were seeded for 48 hours at 1 X 10^6 cells/well in 6-well plates and incubated at 37°C and 5% CO^2. They were then serum-starved for 12 hours in 0.5% FBS. To stimulate early response genes, 20% FBS was added, and cells were collected at 0, 30, 60 and 90 mins. Total RNA was isolated using RNA-tissue kit (Promega) and Maxwell® cDNA synthesis was carried out using high capacity RNA-to-cDNA™ Kit (Thermo Fisher Scientific), followed by TaqMan qPCR analysis (Primer sequence shown in Table 4.1). All experiments were performed with technical replicates and values were normalized using housekeeping genes GAPDH. Post that RNA from each biological replicate was sent to for mRNA sequencing by Novogene Co. Ltd (https://en.novogene.com). The samples were sequenced using the Illumina platform, in a paired end, read length 150 BP and stranded

Box 1:

1. 350000 cell/well were seeded 24 hours prior to transfection in a 6 well plate.
2. 20µM stock of dilute siRNA and negative control were prepared (5.6 µl of siRNA and 14.4 µl buffer blue)
3. Viromer and buffer blue were mixed in the 1:90 µl i.e., 2µl+180µl per sample
4. Next, 180µl of transfection mix of siRNA/Negative control added in tube.
5. Tube was further vertex and incubated for 15 mins at room temperature.
6. Aspirate the old cell-culture and change into fresh one (1800 ml fresh medium)
7. After incubation, add the reaction mixture (200µl) to each cell of 6 well-plate
8. Cells were incubated in 37®c @5% of $C0_2$ for 24 hours.

2.2.1 Bioinformatics methods

i) Datasets used for co-expression analysis

RNAseq expression data for 27 cancer types was downloaded from the FIREHOSE Broad GDAC (https://gdac.broadinstitute.org/), with each cancer dataset having between 60-1000 samples. RNAseq of MDMs (accession number GSE81046) was obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), and the associated raw counts files were downloaded. The dataset included 169 control samples (from healthy individuals). 96 of these were infected *in-vitro* with *Salmonella typhimurium*, and 92 infected *in-vitro* from *Listeria monocytogenes*. In order to avoid problems of stratification, we used only samples of European descent.

ii) Differential expression analysis of MDMs dataset

Genes with < 10 counts across all samples were filtered, raw counts were normalized using the R package `edgeR`, and samples were then grouped based on conditions (controls vs infection states). Differential expression analysis was performed using `limma-voom`, with a cutoff FDR < 0.05,

and a fold change cut of 2.0 between controls and *Salmonella typhimurium*-infected samples, control and *Listeria monocytogenes*-infected samples, as well as *Listeria* monocytogenes-infected and *Salmonella typhimurium*-infected samples (https://github.com/srmeetd/Macrophage_transcriptomic_analysis/blob/main/diff_infection_dataset.R).

iii) Differential expression analysis on time-series data of TRIB1 OE DU145

Data quality of raw reads was checked with `FastQc` tool implemented in the CGAT pipeline readqc (https://github.com/cgat-developers/cgat-flow/blob/master/cgatpipelines/tools/pipeline_readqc.py). Low-quality reads and adapter sequences were removed using the `Trimmomatic` algorithm (Supplementary figure 1). It is worth mentioning that overall data quality was good, but we could see two set batches, based on GC content distribution. Clean reads were then mapped against the human genome hg38, using pseudo-aligner `Salmon`, with parameter "`--gcBias`", which handles GC-biased generating batches in the dataset. Output files generated by Salmon were further passed to `tximport` R package, to retrieve count information. Counts were then normalized using R package `edgeR`, and samples were grouped based on time points, for both TRIB1-overexpressed and control conditions. Differential expression analysis was ultimately performed using `limma-voom` between WT different timepoints vs DU145 TRIB1 overexpressed cell-line. Moreover, we also removed the effect of different timepoints using limma and differential expression was performed between WT and TRIB1 overexpressed DU145 cell-line to counter the effect of TRIB1. And FDR correction cutoff of 0.05 and a log fold change cut off ±1 were applied (https://github.com/srmeetd/Co-expression_analysis_scripts/blob/main/Coexpression_analysis.Rmd).

iv) Steps for generating Co-expression network

mRNA seq raw counts for cancer datasets were downloaded from TCGA Firehose (https://gdac.broadinstitute.org/) and were further processed to generate CPM counts, by using the `limma` R package. The top 5000 highest expressed genes in each cancer dataset and for MDM condition were selected and used to construct the co-expression network for each dataset, using the WGCNA R package. First, `flashClust` R package (included in WGCNA) was used to construct the hierarchical clustering, which helped remove outlier samples from the provided datasets. Next, correlation analysis was performed amongst genes between samples within a dataset, and a default soft threshold was applied to select the positively and negatively correlated genes. Finally, a Topological overlap matrix (TOM) was used as a distance matrix (1-(TOM)), and DynamicTree cut algorithm was used to generate a dendrogram, to cluster highly similar modules and to merge them, with a cutoff of 0.25. Thus, different modules with co-expressed genes were generated through

weighted co-expression network analysis, and we further investigated the module that contained TRIB1 (https://github.com/srmeetd/WGCNA_coexpression).

v) Structural superimposition of TRIB1

In order to test whether there are any structural similarities between Trib1 and an E2 ligase – UBE2D3, their protein structures were downloaded from Protein Data Base (PDB) (PDB Ids 5CEM and 2FUH, respectively) and then superimposed using the online PDB tool.

vi) Gene ontology analysis

Gene ontology enrichment analysis was performed using GOseq (Young et al. 2010) R package. Next, FDR was calculated on the 'over_represented_pvalue' column output obtained from GOSeq. The reason for using the 'over_represented' column is that it represents more DE genes in the present categories than expected at any given size of category and length distribution; hence, these are considered enriched DE genes in that category (Young et al. 2010). Gene ontology passing the threshold of FDR < 0.05 were then selected for further analysis.

2.3 Results

2.3.1 TRIB1 co-expresses with genes of different functions in macrophages

We began our investigation of TRIB1 networks by examining a dataset of monocyte derived macrophages (MDMs) from 169 individuals. As well as control samples (from healthy individuals), 96 samples were infected *in-vitro* with *Salmonella typhimurium*, and 92 samples with *Listeria monocytogenes*. We selected the 5000 most highly expressed genes in each condition, and built three co-expression networks. These had 65 modules for uninfected, 68 for *Salmonella typhimurium*-infected and 62 for *Listeria monocytogenes*-infected macrophage datasets. The modules containing TRIB1 had 107, 57 and 136 genes, respectively (Supplementary Table 1 https://github.com/srmeetd/Coexpression_chapter/blob/main/Supplementary_Table1.xlsx), and we observed that TRIB1 co-expresses with different genes in each condition, with no overlap between the three datasets. (**Figure 2.2**)

As mentioned in Introductory Chapter 1.7.2, TRIB1 is involved in the polarization of M2-like macrophages, and *Listeria monocytogenes* has also been shown to stimulate an M2-like phenotype in macrophages upon infection (Thiriot et al. 2020). Therefore, to better understand TRIB1's role in this type of macrophages, we focused our further analyses on *Listeria monocytogenes*-infected dataset. Thus, out of 136 genes found to co-express with TRIB1 in this condition, 30 were significantly up-regulated and only 2 were down-regulated upon *Listeria monocytogenes* infection, compared to unstimulated MDMs (**Figure 2.3,** Supplementary Table 2 https://github.com/srmeetd/Coexpression_chapter/blob/main/Supplementary_Table2.xlsx).

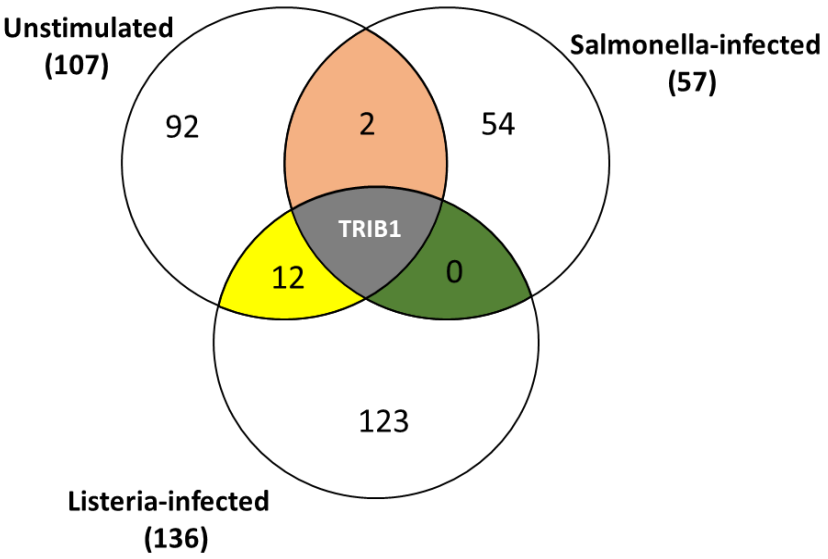


Figure 2. 3- Venn diagram showing number of genes co-expressed with TRIB1 in all conditions

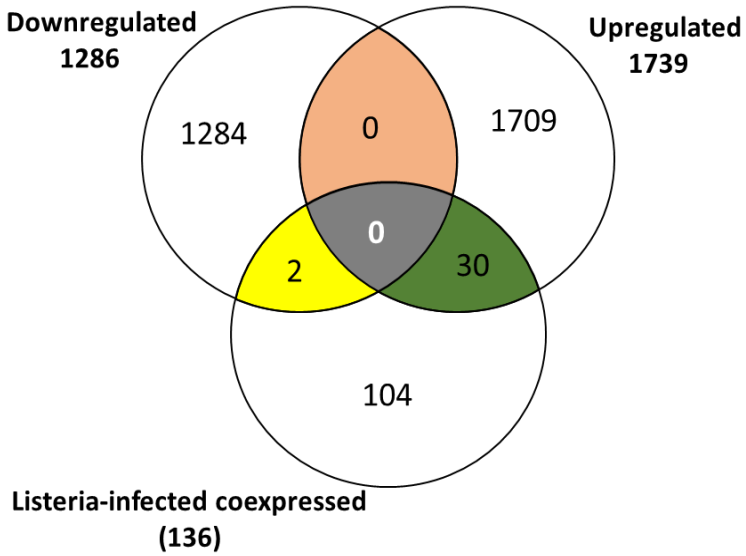


Figure 2. 5- Venn diagram showing genes co-expressed with TRIB1 in Listeria monocytogenes, up-regulated and down-regulated, with fold change ±2 and FDR < 0.05

As discussed in Chapter 1.7.2, TRIB1 has previously been observed to recruit the E3 ligase COP1, for the ubiquitination of CEBPA. Similarly, one of the significantly up-regulated genes co-expressing with TRIB1 in Listeria monocytogenes-infected samples is UNKL (**Table 2.2**), which is also an E3 ubiquitin ligase, playing a role in phagosome formation. This suggests TRIB1 is often involved in the recruitment of E3 ligases. E2 ubiquitin-conjugating enzymes are proteins known to be involved in the recruitment of E3 ligases. One could hypothesise that TRIB1 might act as an E2 ligase. Upon superimposing Trib1 protein structure with that of an E2 ligase, UBE2D3, an overlap of at least 76% was observed (**Figure 2.4**), which supports our hypothesis that Trib1 could have an E2 ligase domain.

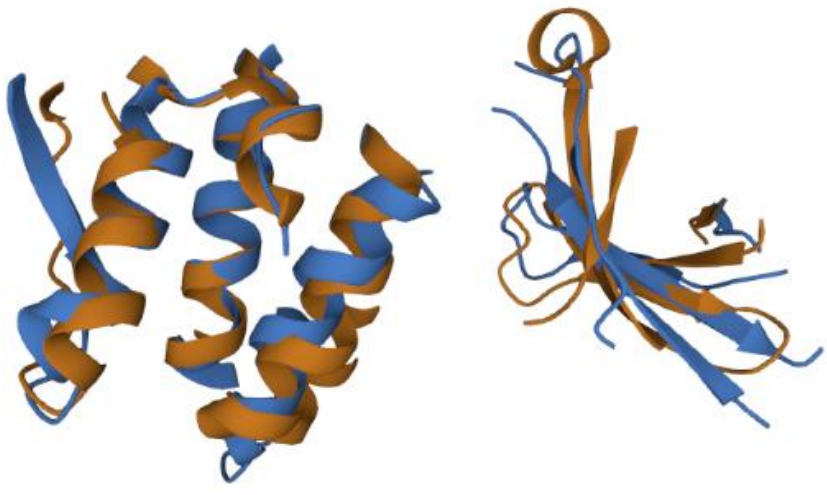


Figure 2. 7- Super-imposed 3D structure TRIB1 (blue) and UBE2D3 (brown)

However, further study is required to understand about the overlapping domains and conclude the significance structural similarities.

Another significantly up-regulated gene co-expressing with TRIB1 is TMEM181 (**Table 2.2**), a transmembrane protein with important role in cell signalling, particularly upon pathogen infections. What is more, looking at the protein network of TMEM181, it is often co-expressed with SYTL3 gene obtained from STRING database (**Figure 2.5**), whose role is to bind calcium ions (Fukuda and Mikoshiba 2001), and together, they are involved in vesicular trafficking and signal transduction. However, *Listeria monocytogenes* is known to sequester calcium ions, which can no longer be bound by SYTL3, thus leading to a lower demand for this protein. This is consistent with our data, as SYTL3 was observed to be down-regulated in *Listeria monocytogenes*-infected samples, suggesting it can no longer ensure a proper signalling with TMEM181. Therefore, the co-expression of TRIB1 with TMEM181 could indicate that Trib1 might mitigate the lower abundance of SYTL3, as well, supporting the signalling pathway through TMEM181 and helping the cells activate their adaptive immunity specific to an M2 phenotype, to fight off the bacterial infection.

The above examples of TRIB1 co-expressing with different types of proteins indicate that Trib1 could have multiple functions, including roles in macrophage polarisation, ubiquitination and cell signalling, thus improving our understanding on TRIB1's potential activity.

Table 2. 2- log fold change of genes co-expressed with TRIB1 in Listeria monocytogenes, either upregulated or downregulated, with fold change threshold ±2

Gene	Log fold change
TRIB1	1.34
UNKL	1.14
TMEM181	0.46
SYTL3	-0.81

2.3.2 TRIB1 co-expresses with immediate-early response (IER) genes

Considering TRIB1 is known to be upregulated in many different types of cancer (see Introduction 1.7.2 ii)), we generated co-expression networks of TRIB1 for different cancer datasets, and then looked at whether the genes identified to co-express with TRIB1 in any of these datasets were overlapping with those detected in macrophages. Thus, we obtained a separate set of modules

for each cancer type, containing between 20-80 genes co-expressed with TRIB1 per cancer type. In order to narrow down the list of potential interactomes, we selected only the genes co-expressing with TRIB1 in at least eight cancer datasets, finding only 28 such genes (columns in **Figure 2.6**).

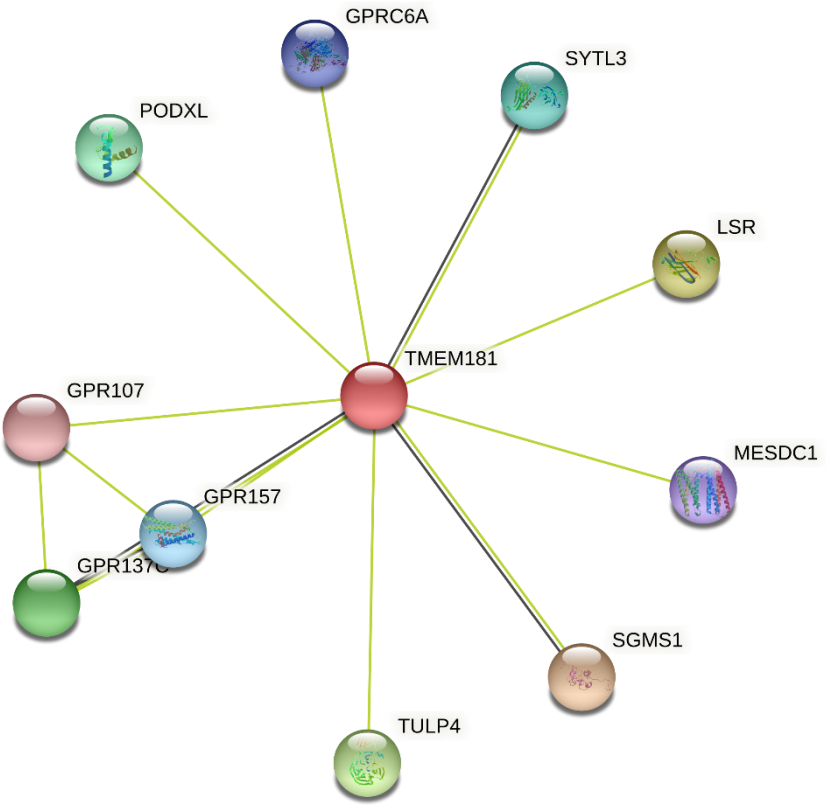


Figure 2. 9- Interaction network of proteins, obtained from genes co-expressed with TMEM181; yellow – text-mining evidence; black – co-expression evidence

These 28 genes were coexpressed with TRIB1 in upto 18 different cancer types, and in general TRIB1 would either co-expressed with many or none of these genes in a given cancer type. In 9 out of the 28 cancer datasets analysed, TRIB1 co-expressed with few or none of these genes, but instead a different set of genes were co-expressed with TRIB1 (rows in **Figure 2.6**). Out of the 28 genes mentioned above (columns in **Figure 2.6**), 19 are known to be immediate-early response genes (IER) (**Table 2.3**). These IER genes were not co-expressed in the 9 outlier cancer datasets (rows in **Figure 2.6**). These 17 IER genes were lowly expressed in these 9 cancer data sets (in which they did not co-express with TRIB1), compared to the other 21 data sets (in which they were co-expressed with TRIB1) (**Figure 2.7,** Wilcox test p-value = 0.02). In contrast, TRIB1's expression was not significantly different between samples in which IER genes were co-expressed with it and those in which they weren't (**Figure 2.8**, Wilcox test p-value = 0.96). These results show that TRIB1 is co-expressed with a high number of IER genes in 19 out of 28 cancer data sets, suggesting it could potentially be an IER gene itself, regulate an IER gene or be regulated by an IER gene.

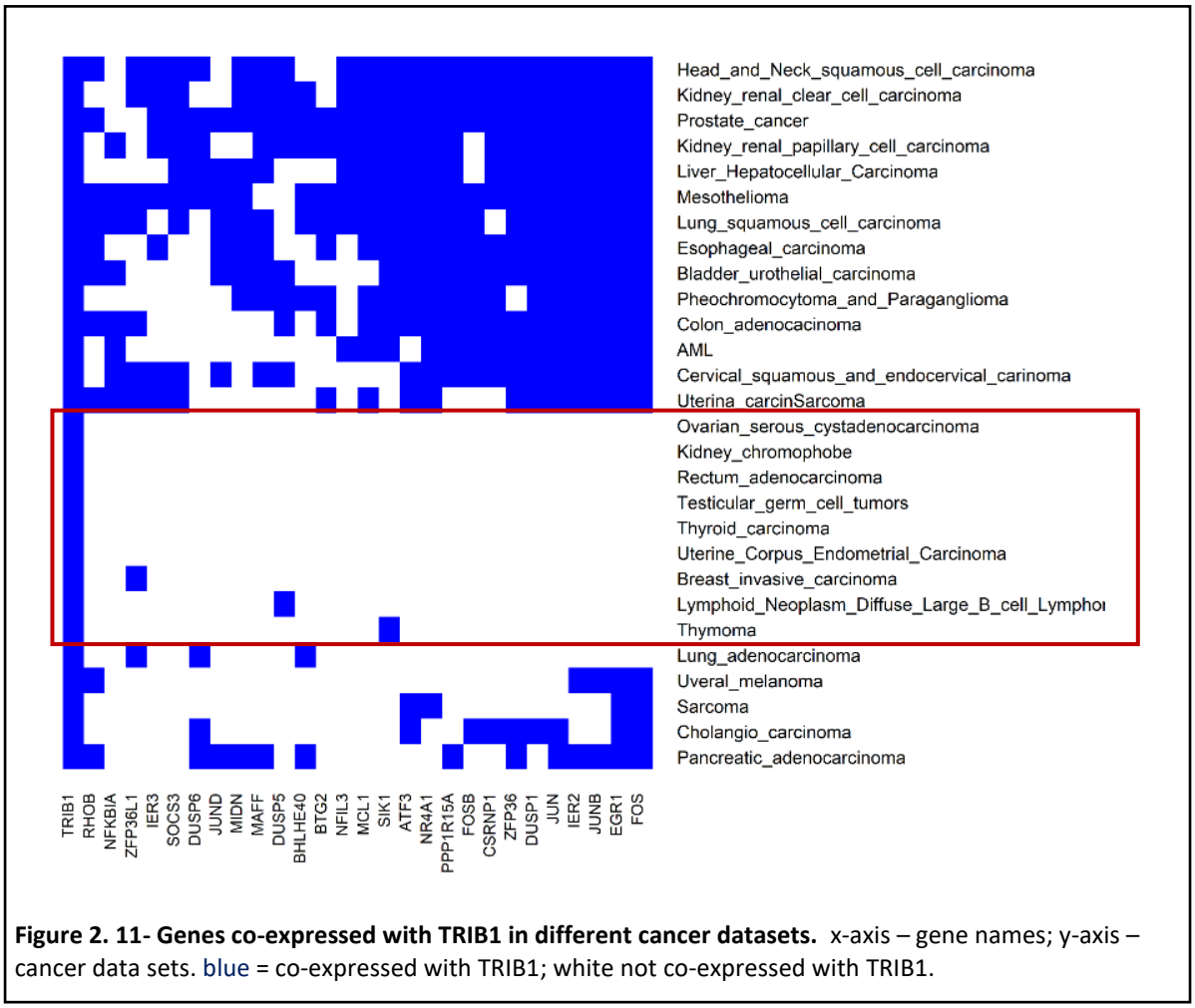


Figure 2. 11- Genes co-expressed with TRIB1 in different cancer datasets. x-axis – gene names; y-axis – cancer data sets. blue = co-expressed with TRIB1; white not co-expressed with TRIB1.

Table 2. 3- List of early response genes co-expressed with TRIB1

Genes	No. of cancer datasets	References
EGR1	19	(Duclot and Kabbaj 2017)
FOS	19	(Sng, Taniura, and Yoneda 2004)
IER2	17	(Neeb *et al.* 2012)
JUN	17	(Hernandez *et al.* 2008)
JUNB	17	(Hernandez *et al.* 2008)
ATF3	17	(Epanchintsev *et al.* 2020)
DUSP1	16	(Horita *et al.* 2010)
NR4A1	16	(Tullai *et al.* 2007)
PPP1R15A	16	(Uhlitz *et al.* 2017)
ZFP36	16	(Uhlitz *et al.* 2017)
FOSB	14	(Tullai *et al.* 2007)

Table 2. 4- List of early response genes co-expressed with TRIB1

Genes	No. of cancer datasets	References
MAFF	13	(Kannan, Solovieva, and Blank 2012)
MCL1	13	(Tullai *et al.* 2007)
DUSP	11	(Horita *et al.* 2010)
JUN	11	(Hernandez *et al.* 2008)
RHOB	11	(Huelsenbeck *et al.* 2007)
ZFP36L	10	(Uhlitz *et al.* 2017)
DUSP5	8	(Horita *et al.* 2010)
DUSP6	8	(Horita *et al.* 2010)

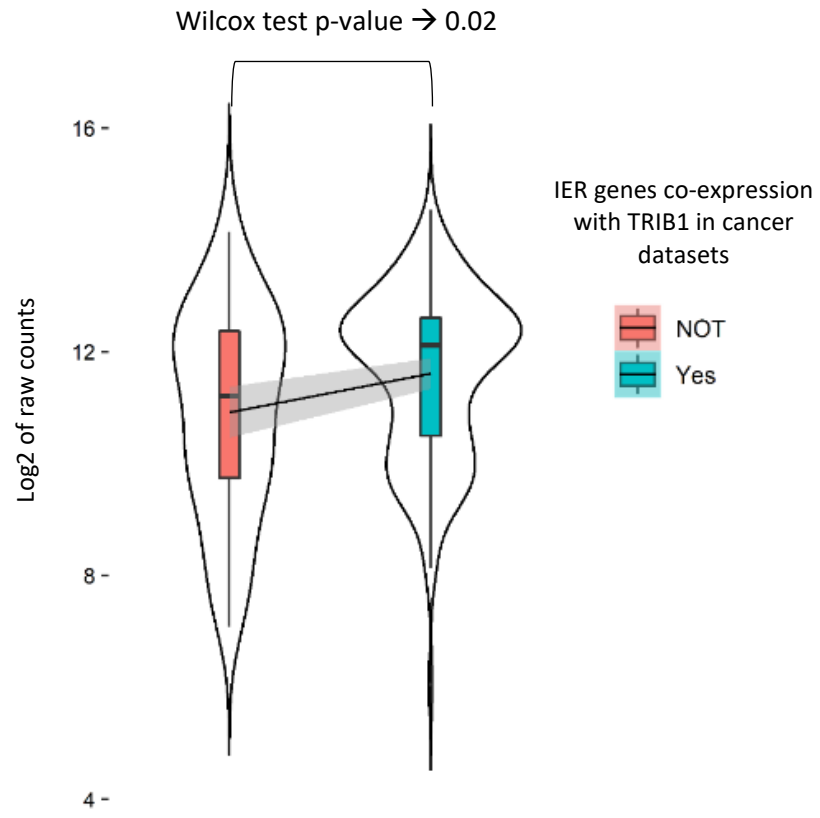


Figure 2. 13- Box-violin plot showing the expression of IER genes co-expressed (green) or not (orange) with TRIB1 in cancer datasets

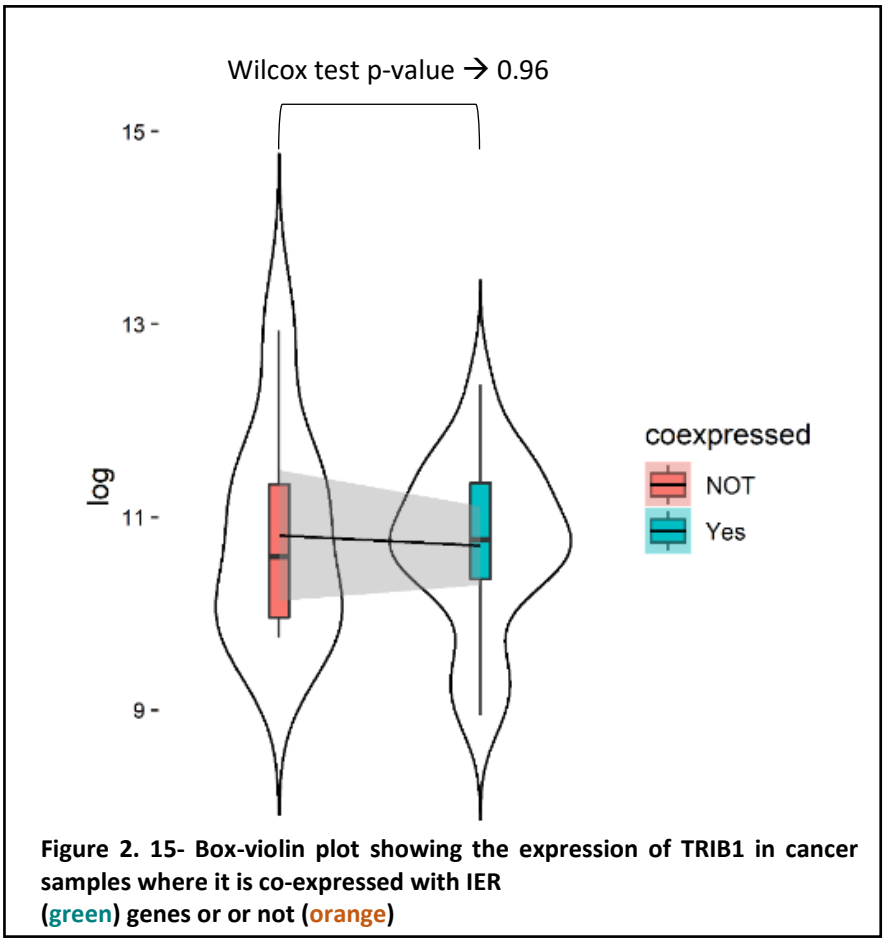


Figure 2. 15- Box-violin plot showing the expression of TRIB1 in cancer samples where it is co-expressed with IER (green) genes or or not (orange)

WGCNA generated modules can contain genes either positively or negatively correlated with each other between samples. We examined the Pearson correlation between genes in the TRIB1 module to observe the direction of correlation between IER genes and TRIB1. We found that IER genes were positively correlated with TRIB1 in ~90% of cancer datasets (**Figure 2.9**), meaning that when the expression levels of either TRIB1 or these IER genes go up/down then the expression of the others also goes up/down.

2.3.3 TRIB1 co-expresses with genes involved in different pathways in cancer

As mentioned before, TRIB1 has been associated with a number of signalling pathways and biological processes involved in cancer, but its exact role is yet to be elucidated. In order to better understand the network interactions of TRIB1, we investigated the genes co-expressed with it in both *Listeria monocytogenes*-infected and cancer samples. For this, we used the GOseq and KEGGREST R package for KEGG-pathway enrichment analysis, and we observed that genes co-expressed with TRIB1 in *Listeria monocytogenes* dataset were not significantly enriched for any KEGG pathways. In contrast, genes recorded in co-expression modules with TRIB1 for each different cancer type were found to be involved in MAPK signalling pathways in 20 out of 23 cancer datasets, which was consistent with previously-published findings of TRIB1 involved activating MAPK pathways (Zhang et

al. 2021). Other well-supported pathways were B cell receptors, followed by cytokine, T-cell signaling pathways, and Toll-like receptor pathways (**Figure 2.10**). Therefore, for further validation, we focused on analysing genes co-expressed with TRIB1 in cancer.

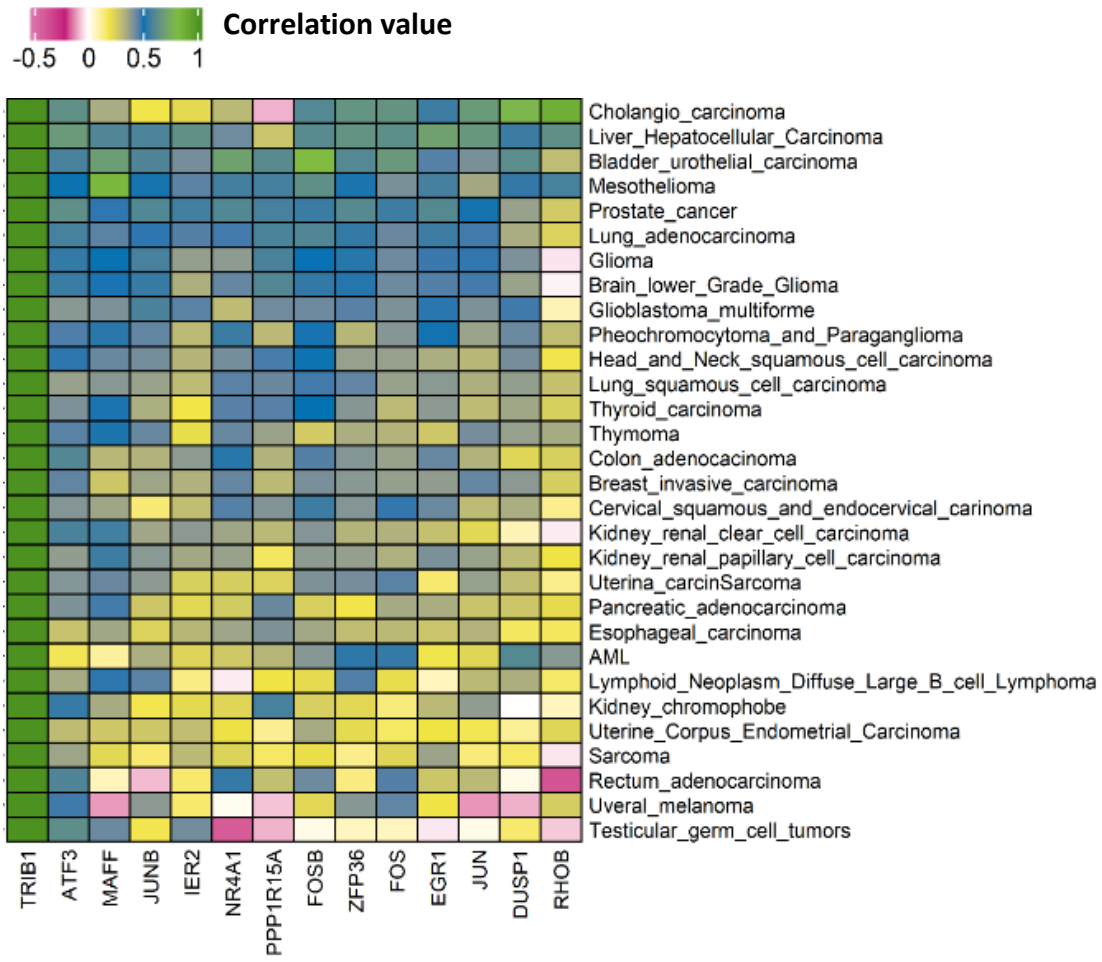
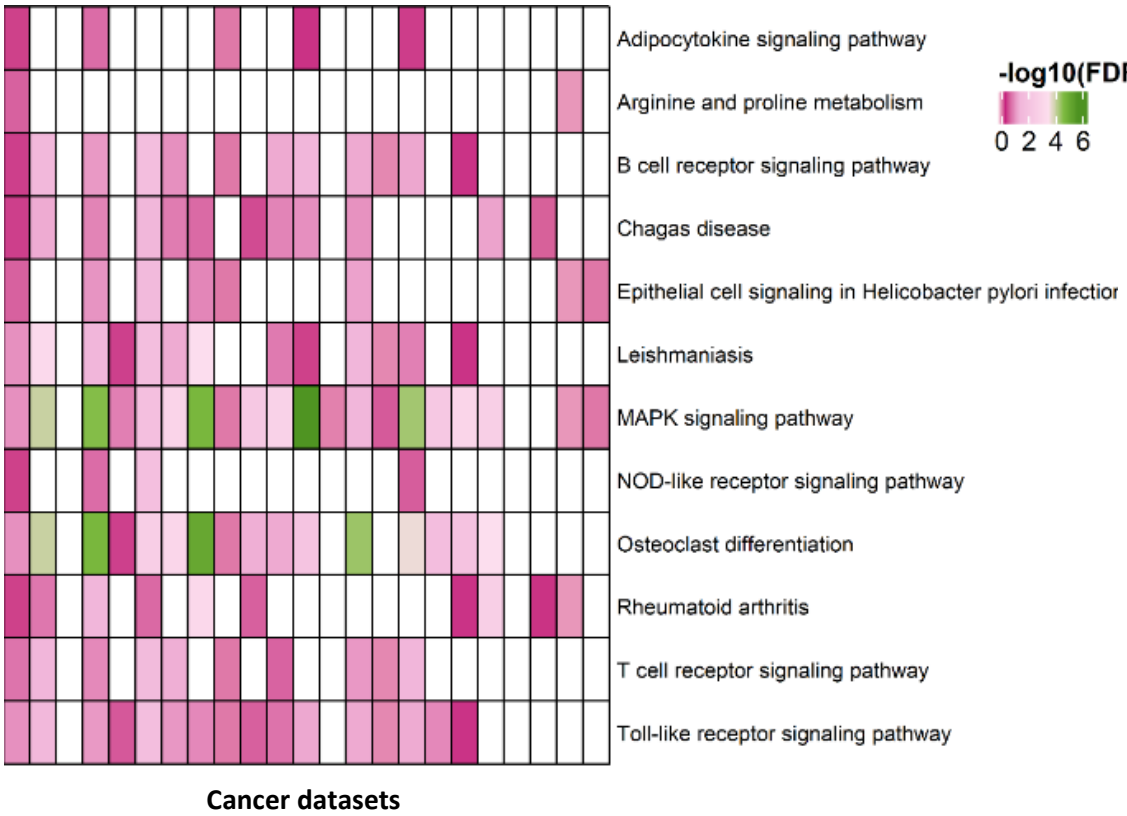


Figure 2. 17- Correlation of IER genes with TRIB1 in different cancer types

2.3.4 TRIB1 co-expresses and interacts with IER genes EGR1 and FOS

The protein-protein interactions of genes co-expressing with TRIB1 in more than ten cancer datasets were analysed using the STRING database. The network generated recorded a larger number of edges than the expected random network of the same size (Szklarczyk *et al.* 2021), with a p-value ≤ 10^{-16}, which meant the nodes were connected and that the interactions observed were not at random. Overall, only nine protein-protein interactions with Trib1 were identified in STRING, among which the Fos, Dusp1, and Junb proteins, also found in the co-expression network, were found to interact with Trib1 directly (**Figure 2.11**).



2. 21- Heatmap representing top 10 pathways of co-expressed genes present in different 23 cancers

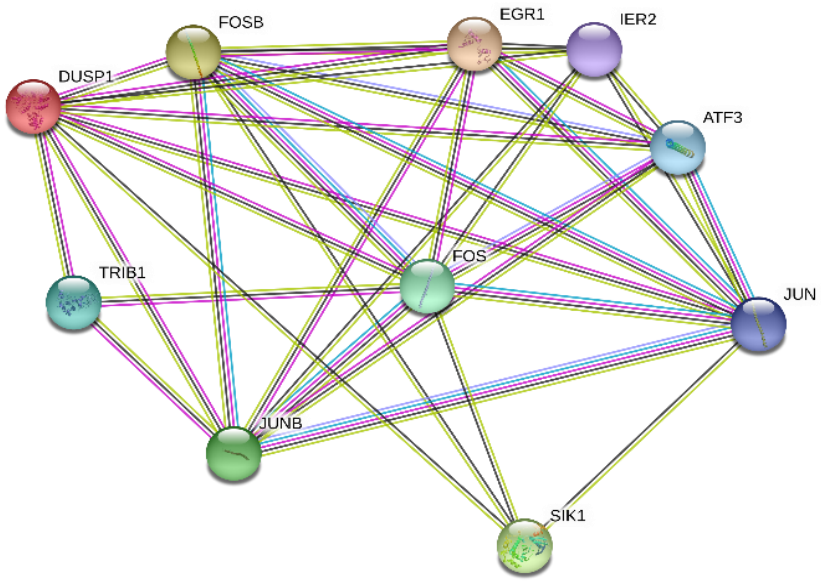


Figure 2. 19- Protein-protein interaction network of genes co-expressed with TRIB1 in cancer datasets. purple – experimental evidence; green – gene neighbourhood; blue – gene co-occurrence database evidence; yellow – text-mining evidence; **black** – co-expression evidence.

2.3.6 TRIB1 has an IER expression pattern in HEK293T cells

In order to test our hypothesis that TRIB1 may be an early response gene, we examined its expression pattern in HEK293T cells in which early response was stimulated through serum supplementation after starvation (The reason to use this method was to restrict the cell growth to G0 phase and create a cell-stress like condition and adding the fresh serum activates mitogenic growth factors responsible for cell proliferation). We observed that TRIB1 expression increased for up to 90 minutes post-stimulation with 20% FBS and then gradually decreased (**Figure 2.12**, Supplementary Table 3 https://github.com/srmeetd/Coexpression_chapter/blob/main/Supplementary_Table3.xlsx). This was consistent with the pattern of known IER genes like EGR1 and FOS, which were also shown to co-express with TRIB1. Although the difference in expression between the time points was not as significant as for EGR1 and FOS, these results support our hypothesis and the possibility of an interaction between these genes.

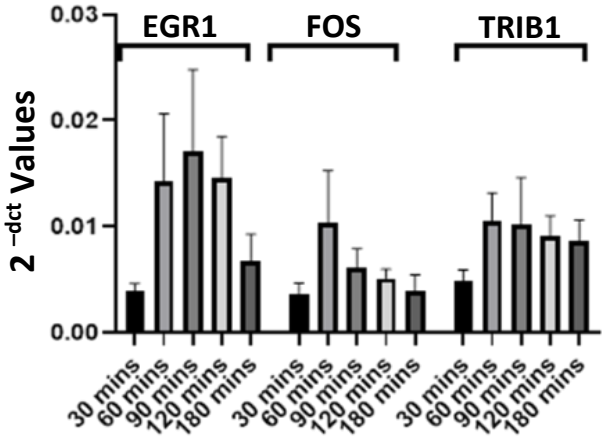


Figure 2. 23- qPCR - mRNA expression of TRIB1, EGR1 and FOS in HEK293T cells, normalized to GAPDH. Data is represented as mean ± SEM (3 bioreps).

2.3.5 Expression of TRIB1, EGR1 and FOS in PC-3 and DU145 cell-line

As mentioned in the Introductory Chapter 1, TRIB1 has been associated with different types of cancer, particularly with prostate cancer pathogenesis, and it was also reported to be over-expressed in prostate cancer cells (Niespolo et al. 2020). The two IER genes we demonstrated above to be co-expressed with TRIB1, EGR1 and FOS, have also been shown to be involved in prostate cancer: EGR1 is a known to trigger apoptosis in prostate cancer cells, by increasing the expression of tumour necrosis factors (Gitenay and Baron 2009); FOS has roles in cell growth and apoptosis, and its expression levels were found elevated in prostate cancer, as well (Shankar et al. 2016). Considering such past studies were performed on both samples from cancer patients and prostate cancer cell lines (e.g. PC-3), in order to test our preliminary computational results on the relationship between

TRIB1, EGR1 and FOS, we also conducted further *in vitro* experiments on prostate cancer cell lines PC3 and DU145.

In order to further understand the possible relationship between TRIB1, EGR1 and FOS mRNA expression, we analysed the mRNA expression data for both DU145 and PC3 cell lines from the CCLE database of gene expression in cancer cell lines. Plotting strip charts for the CPM counts of TRIB1, EGR1 and FOS, we observed that TRIB1 is higher in PC-3 cells than in DU145, while the opposite was observed for both EGR1 and FOS, whose expression levels were elevated in DU145, compared to PC-3 cells (**Figure 2.13**). These indicate that TRIB1 may actually be lowly expressed in DU145 compared to PC3 cell-line. Also, the expression of FOS is more in DU145 compared to PC3, but the expression of EGR1 is slightly higher in DU145 compared to PC3. This data also suggests the possible relationship between TRIB1 , EGR1 and FOS.

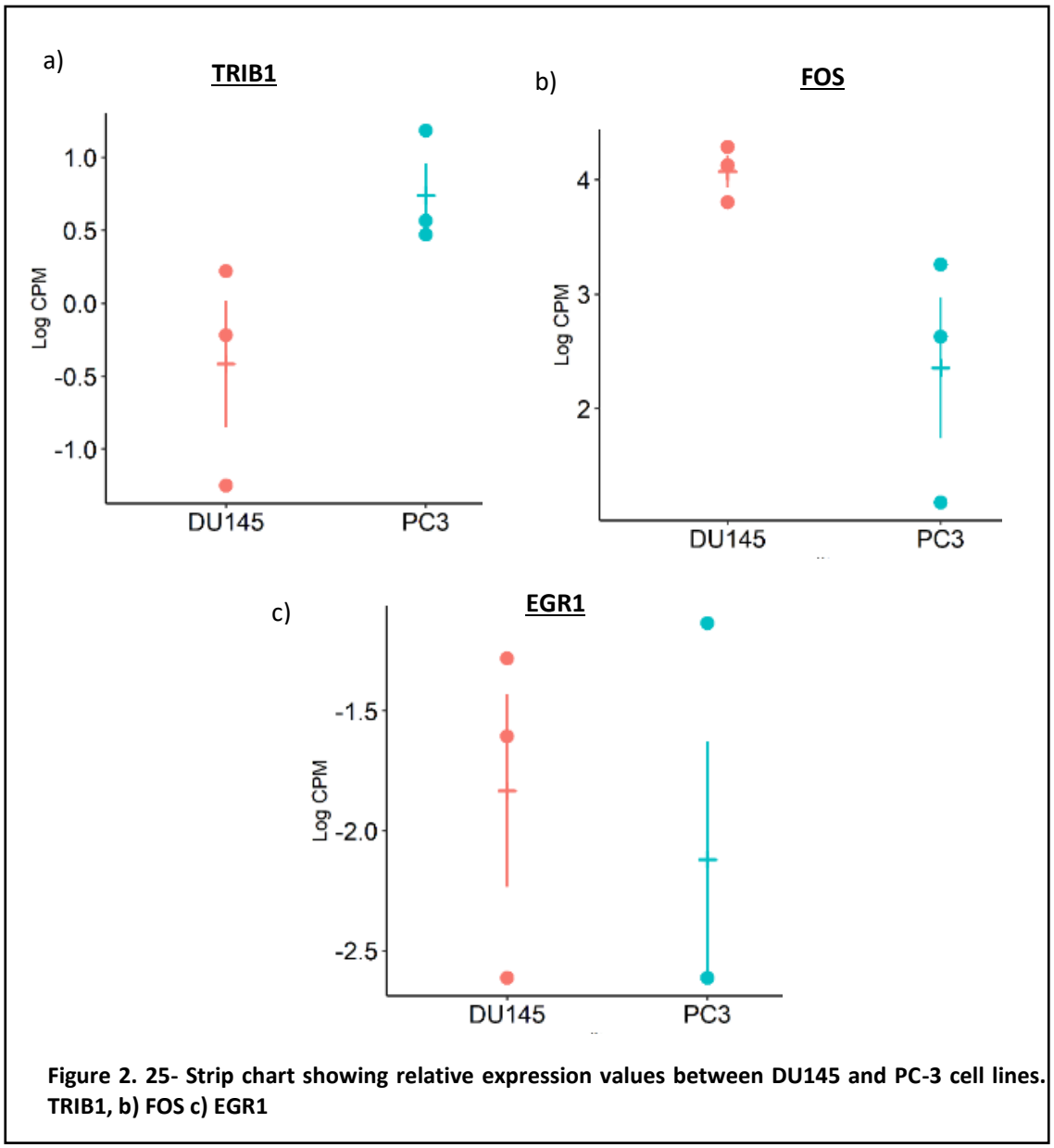


Figure 2. 25- Strip chart showing relative expression values between DU145 and PC-3 cell lines.) TRIB1, b) FOS c) EGR1

2.3.6 TRIB1 may regulate early response of EGR1 and FOS

We tested this connection between TRIB1, EGR1 and FOS, by overexpressing TRIB1 in DU145 cells and knocking it down in PC-3, as well as in HEK293T cells. qPCR analysis results showed an obvious increase in EGR1 expression in TRIB1 knockdown conditions in both PC-3 and HEK293T cell lines, while FOS levels were only slightly higher (**Figure 2.14,** Supplementary Table 4 https://github.com/srmeetd/Coexpression_chapter/blob/main/Supplementary_Table4.xlsx). In agreement with **Figure 2.13,** this would suggest TRIB1 could inhibit IER genes like EGR1. If so, it would be expected that in cells overexpressing TRIB1, EGR1 levels would significantly decrease. However, when TRIB1 was overexpressed in DU145 cells, both EGR1 and FOS were significantly upregulated (**Figure 2.14**), contrary to our initial predictions. One possible reason for this could be that overexpression of TRIB1 in DU145 cells (which in WT conditions it was observed to be lower in DU145 than in PC-3 cells – **Figure 2.13**) could have resulted in TRIB1 levels similar to those recorded when it was knocked down in HEK293T and PC-3 cells. This suggests there could be a particular threshold or cellular demand for TRIB1 that needs to be maintained and when dysregulated, IER genes are

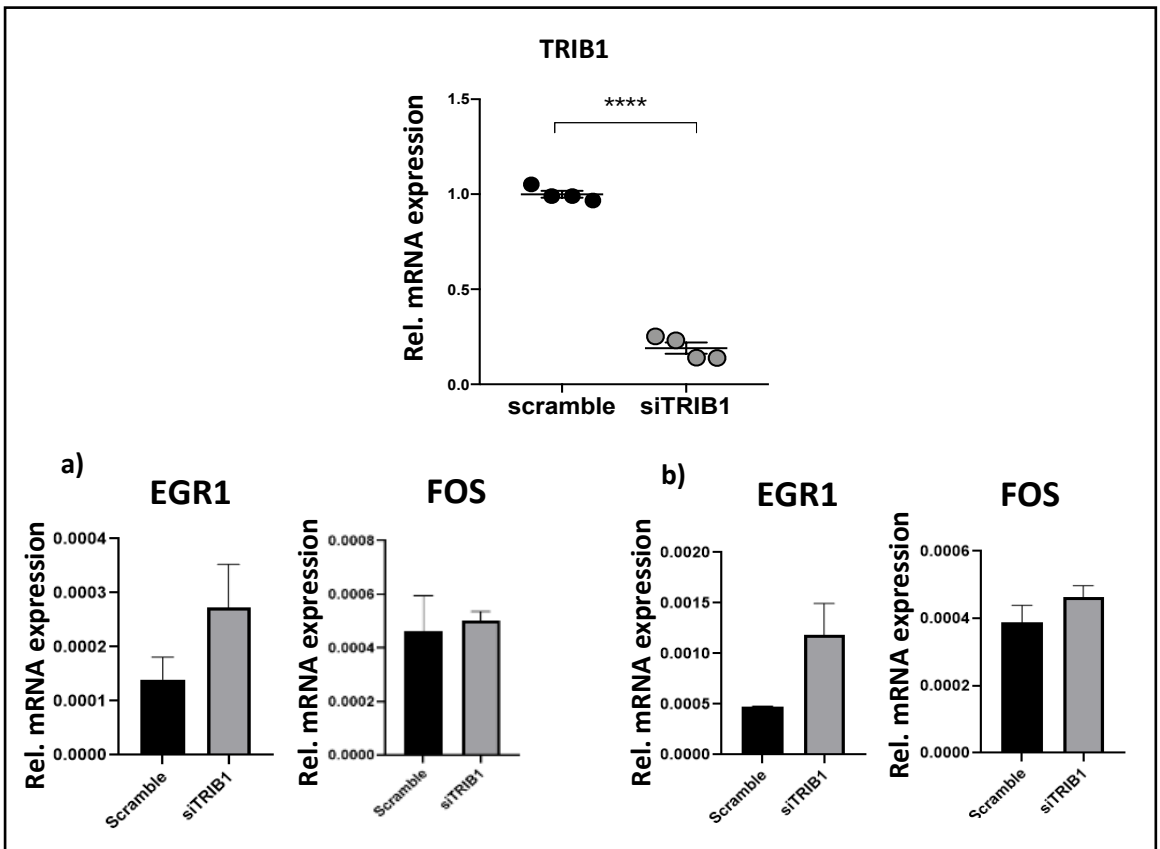


Figure 2. 27- qPCR - mRNA expression of EGR1 and FOS in TRIB1 knock down in a) PC-3 and b) HEK293T cells, normalized to the housekeeping gene GAPDH. Data is represented as mean ± SEM (3 tech reps). TRIB1's expression was only measured in PC-3 cells (top).

positively stimulated. However, it is important to note that because of the COVID-19 related lockdowns, the data in **Figure 2.14** represents only a single biological replicate.

2.3.6 Change in expression level of TRIB1 upon stimulating IER genes

The above experiments took place in resting cells. In order to test whether TRIB1 has any effect on activation of EGR1 and FOS in conditions triggering early response, the abundance of both genes was analysed at different time points of serum supplementation of DU145 cells overexpressing TRIB1. qPCR analysis results showed a significant increase in EGR1 expression at 30 mins in TRIB1 overexpressed cell-lines (**Figure 2.16**), similar to that recorded in DU145 overexpressing TRIB1 grown in normal media (**Figure 2.15**). However, EGR1 levels started to decrease comparative to WT DU145 cells after 90 minutes post-stimulation (**Figure 2.16**), suggesting TRIB1's overexpression could have an early impact on EGR1. In contrast, although FOS levels were recorded to be higher in DU145 cells overexpressing TRIB1 and grown in normal media, when early response was serum stimulated in the same cells, the expression of FOS was drastically reduced at and after 30 mins in TRIB1 overexpression cell-line (**Figure 2.16**), again indicating that TRIB1 may have an early effect on these genes.

Further overexpression studies in HEK293T and PC-3 cells, as well as knockdown of TRIB1 in DU145 cells would help elucidate whether this behaviour is cell line specific and whether there is a direct effect of one gene over the other(s).

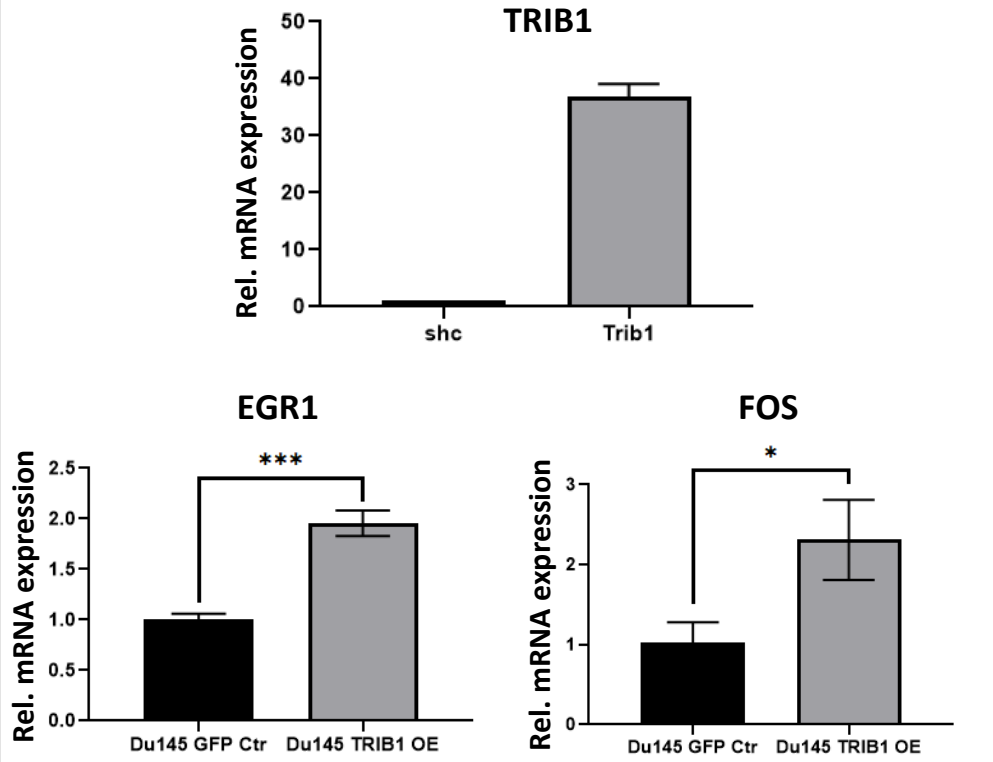


Figure 2. 29- mRNA expression of TRIB1, EGR1 and FOS in DU145 cells over-expressing TRIB1 (Data is represented as mean ± SEM).

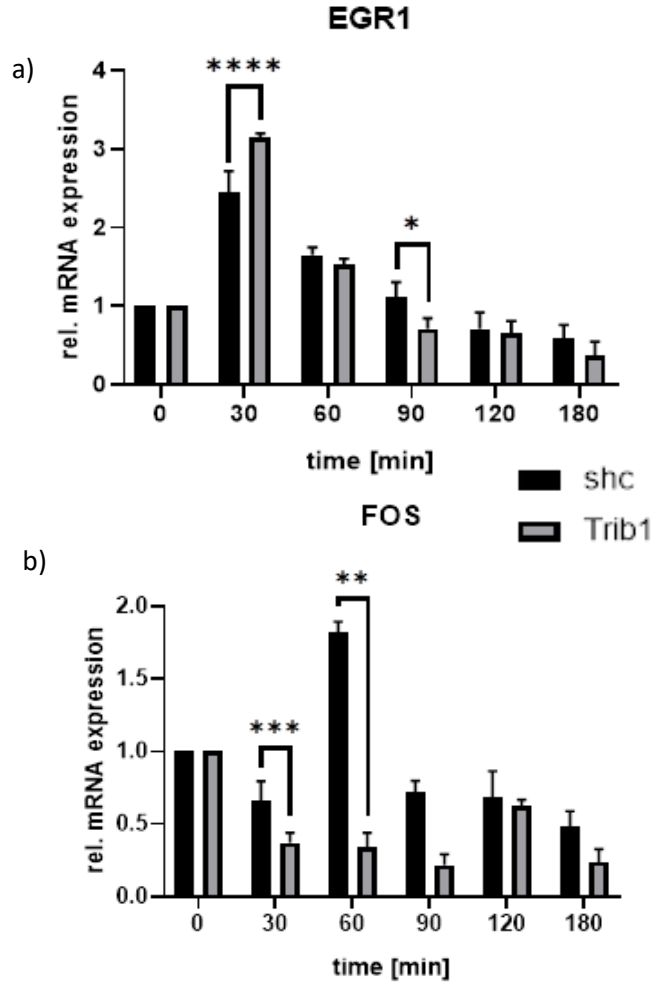


Figure 2. 31- : mRNA expression of a) EGR1 and b) FOS at different time-points in DU145 cells overexpressing TRIB1, normalized to the housekeeping gene GAPDH **Data is represented as mean ± SEM (3 biological replicate) (Shc → Scrambled control)**

2.3.6 Expression of TRIB1 is negatively correlated with that of EGR1 and FOS in DU145 and PC-3 cell lines

We tested the abundance of TRIB1 in DU145 cells at different time points of post-starvation stimulation with 20% FBS, which replicated the conditions in which IER genes would be expressed. qPCR analysis was performed on two different sets of samples, one collected at 6 time points (0, 30, 60, 90, 120 and 180 mins), and the other at only 4 time points (0, 30, 60 and 90 mins). The qPCR results showed a similar relative abundance of TRIB1 mRNA between the different time points in shc control cell lines with increase in TRIB1 expression at 90 minutes. However, in TRIB1 OE DU145 cell lines TRIB1 at different time-points, we do not observe the much variation. (**Figure 2.17 and 2.18**). This could be because TRIB1 is already OE in DU145 cell-line, hence it is not further regulated by IER

genes. Moreover, we observed the different in expression level in **Figure 2.17 and 2.18**, although the direction of expression is similar. These could be due to a number of factors, such as technical error, efficiency of qPCR primers, differences in detection/sensitivity of the machine on different days etc. Despite this, it can be observed there was small amount of increase in the expression levels of TRIB1 between the different time points, suggesting TRIB1 may not in fact act like an early response gene in DU145 prostate cancer cells. The most possible reason could be TRIB1 is overexpressed in DU145 cell lines therefore it is not further activated after stimulating IER genes.

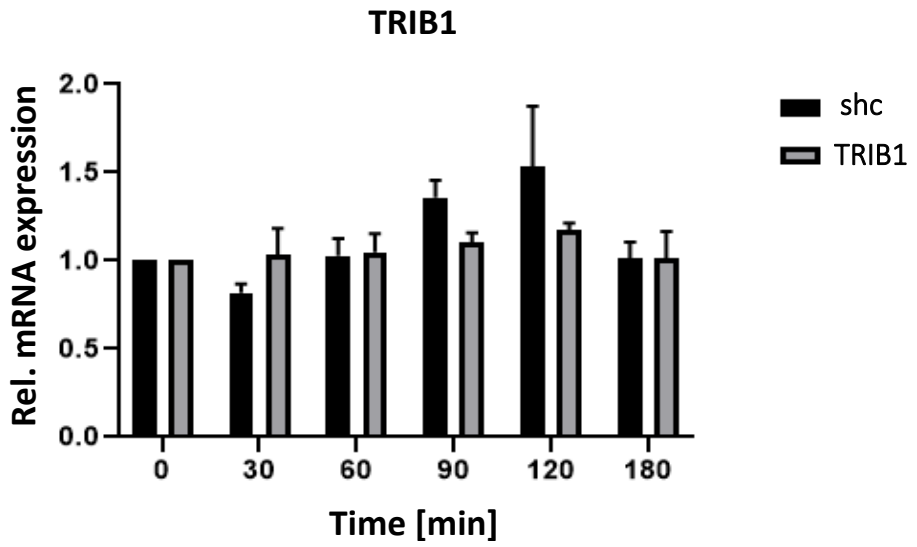


Figure 2. 33- qPCR - mRNA expression of TRIB1 in DU145 cells at 0, 30, 60, 90, 120 and 180 mins post-starvation for 12 hours and stimulation with 20% FBS (Data is represented as mean ± SEM) (Shc → Scrambled control)

2.3.8 TRIB1 overexpression leads to differential expression of many genes

In order to validate our computational results obtained from analysing samples available on online databases, as well as our experimental findings, WT DU145 cells and DU145 cells

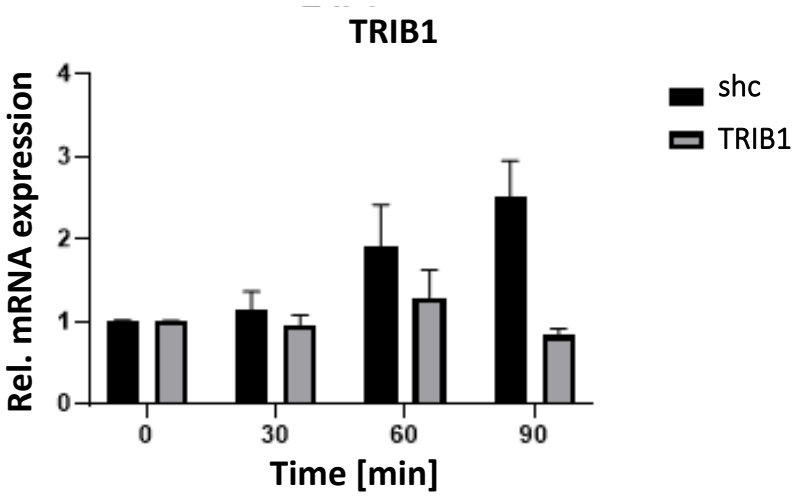


Figure 2. 35- mRNA expression of TRIB1 in DU145 cells at 0, 30, 60 and 90 mins post-starvation for 12 hours and stimulation with 20% FBS (Data is represented as mean ± SEM). (shc → Scrambled control, TRIB1 – TRIB1 OE).

overexpressing TRIB1 were collected at different time points after serum stimulation and sent for mRNA sequencing. The same conditions were described in section 4.3.6, but the RNA amount left from those initial tests was not sufficient for RNAseq. Thus, new cells were thawed and a different batch of DMEM media was used to replicate the previous set of experiments. The expression of EGR1 and FOS was again tested using qPCR (**Figure 2.19**), and the results obtained were different than those reported above in **Figure 2.16**. Although we could not confirm whether the new batch of cells or media were responsible for the different results, the samples were still sent for sequencing.

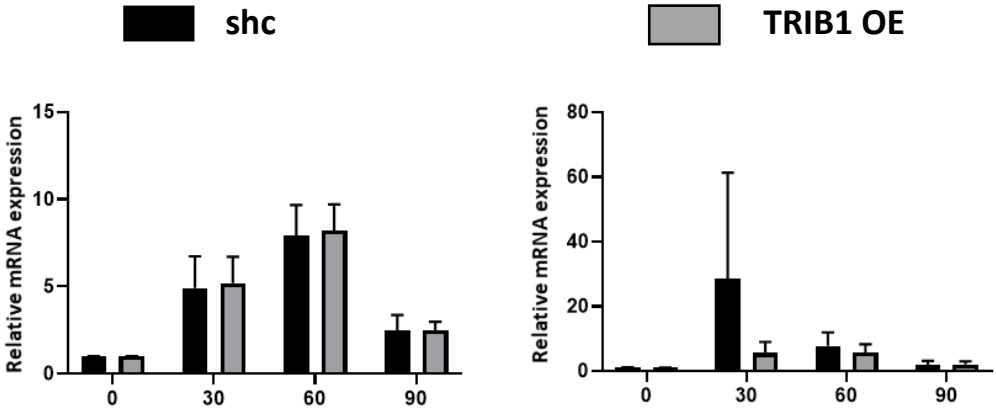


Figure 2. 37- qPCR - mRNA expression of a) EGR1 and b) FOS at different time-points in DU145 cells overexpressing TRIB1, normalized to the housekeeping gene GAPDH. Data is represented as mean ± SEM (3 biological replicate)

Alignment of the RNAseq data resulted in approximately 80% of reads mapping to the human genome hg38 (**Appendix table 7**). Downstream analysis of the rlog values calculated through DESEQ2 R package was visualized through a PCA plot (**Figure 2.20**), in which we could observe that samples were separated by both time points (0, 30 and 60 minutes – principal component 1 - PC1) and condition (control vs. TRIB1 overexpression – principal component 2 – PC2).

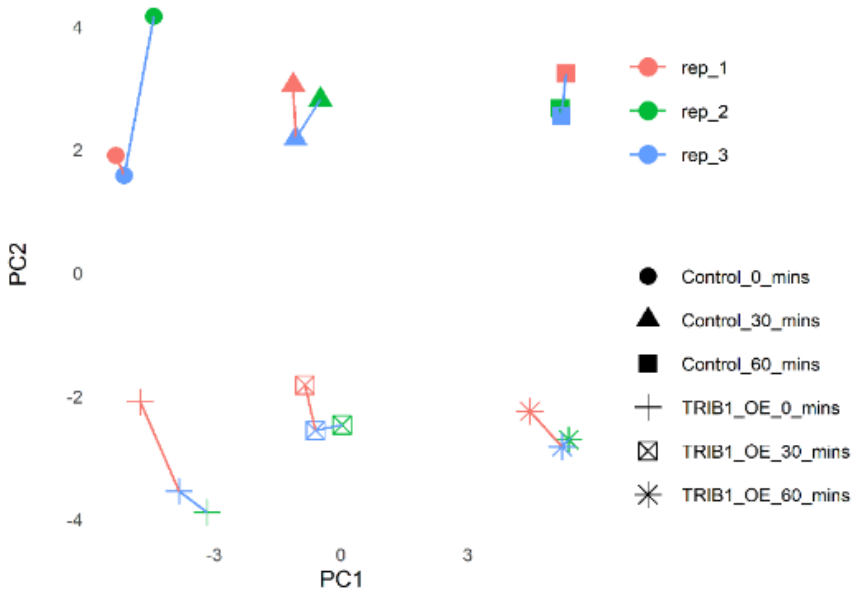


Figure 2. 39- Global PCA plot based on rlog values from RNAseq, performed on GFP control and TRIB1-overexpressed mRNA. Different colours represent the biological replicates, and shapes represent time points in both TRIB1 and control samples.

To investigate the effect of TRIB1 on early response gene stimulation in DU145 cells at different time points, we used DESeq2 to perform differential expression analysis between control DU145 and cells with TRIB1 OE (Supplementary Table 5 https://github.com/srmeetd/Coexpression_chapter/blob/main/Supplementary_Table5.xlsx). We found 85 differentially expressed genes (FDR < 0.05) at time point 0, out of which 48 were upregulated and 37 were downregulated, followed by 30 genes (15 upregulated and 15 downregulated) at time point 30 min, and 21 (12 upregulated and 9 downregulated) at 60 minutes post-stimulation. However, none of the genes identified belonged to the early response genes class. Furthermore, to understand the overall effect of TRIB1 overexpression, we performed differential expression analysis between DU145 controls and cells with TRIB1 OE, irrespective of their time points, accounting for the replicate. As a result, we found 1214 differentially expressed genes (FDR < 0.05), out of which 638 were upregulated and 576 downregulated.

As shown in **Figure 2.21**, there is a clear clustering of genes differentially expressed between DU145 controls and cells with TRIB1 OE, regardless of the 3 time points post-stimulation. Overall,

there is a larger number (638 upregulated and 576 downregulated) of genes being upregulated than downregulated in TRIB1 OE condition, compared to controls, suggesting a positive regulatory effect of TRIB1.

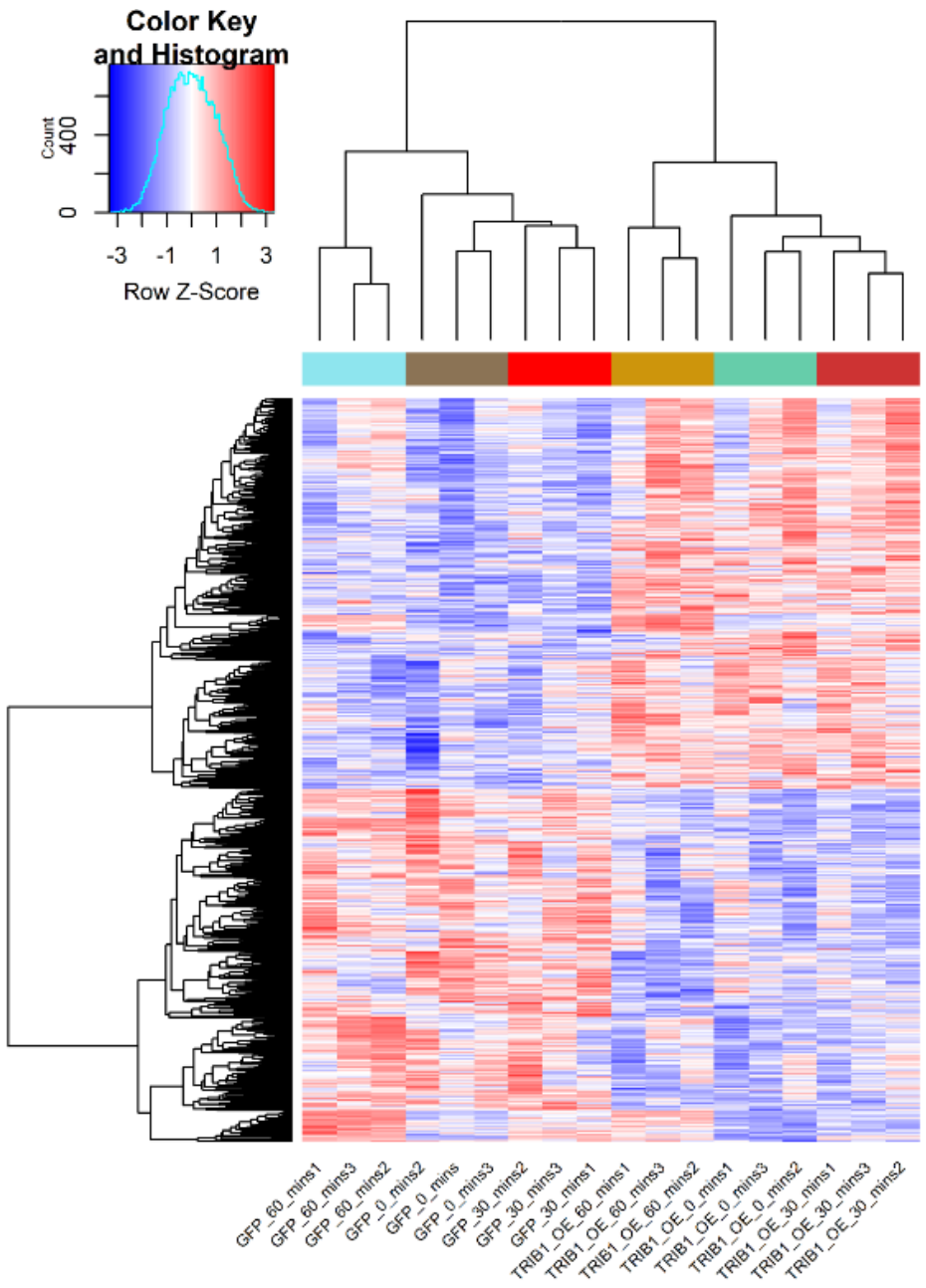


Figure 2. 41- Heatmap of differentially expressed miRNAs in TRIB1-overexpressed and control samples, at different time points. Heatmap is generated from the distance matrix on Pearson correlation, using rlog normalized values

2.3.9 TRIB1 co-expresses with DE IER genes involved in cell signalling and cell cycle related pathways

Although there were a large number of genes differentially expressed between DU145 controls and cells with TRIB1 OE, there was only a small overlap between genes differentially expressed overall and at different time points post-stimulation (only 100 common genes). Therefore, in order to understand whether these genes were involved in different biological processes, gene ontology enrichment was applied. The analysis showed that most of the genes differentially expressed between controls and TRIB1 OE, irrespective of the time points, were involved in various metabolic and catabolic processes (**Figure 2.22a**). In contrast, genes with significantly altered expression at individual time points were associated with cell signalling and cell-related biological functions, such as cell migration, mobility and differentiation (**Figure 2.22b**).

In order to better understand the potential interaction of TRIB1 with these DE genes in prostate cancer, the genes recorded as differentially expressed between DU145 controls and cells with TRIB1 OE were compared with the genes identified through co-expression analysis of RNAseq data for DU145 cells. Thus, we observed that 18 out of 80 genes co-expressing with TRIB1 in prostate cancer were also differentially expressed between DU145 controls and TRIB1 OE. Most of these genes were found to be early response genes involved in cell signalling-related pathways, including BCL6, CLDN4, ELF3, FURN, GADD45B, VEGFA, PHLDA1, RASD1 and TSC22D1 (**Figure 2.23**). In addition, aiming to identify the proportion of differentially expressed genes also co-expressing with TRIB1 in different cancer datasets, we found that 14 out of 28 genes co-expressed with TRIB1 in more than five cancer datasets were also differentially expressed between DU145 controls and TRIB1 OE, at different time points. These are DUSP1, DUSP5, DUSP6, JUN, JUNB, MCL1, PPPP1R15A, RHOB, NFKBIZ,AG1, GPX3, TNFAIP3, RRAD and SNHG1 (**Figure 2.24**), all early response genes associated with cell signalling and cell cycle-related pathways, similar to the ones identified in prostate cancer cell lines.

These results support previous reports of TRIB1 being involved in processes linked to cell fate, and they suggest that TRIB1 may indeed be an early response gene or in direct/close interaction with early response genes. Moreover, these findings strongly indicate that TRIB1 is involved in cell signalling, in line with our above reports of TRIB1 co-expressing with transmembrane protein TMEM181, also known to regulate cell signalling processes upon bacterial infections.

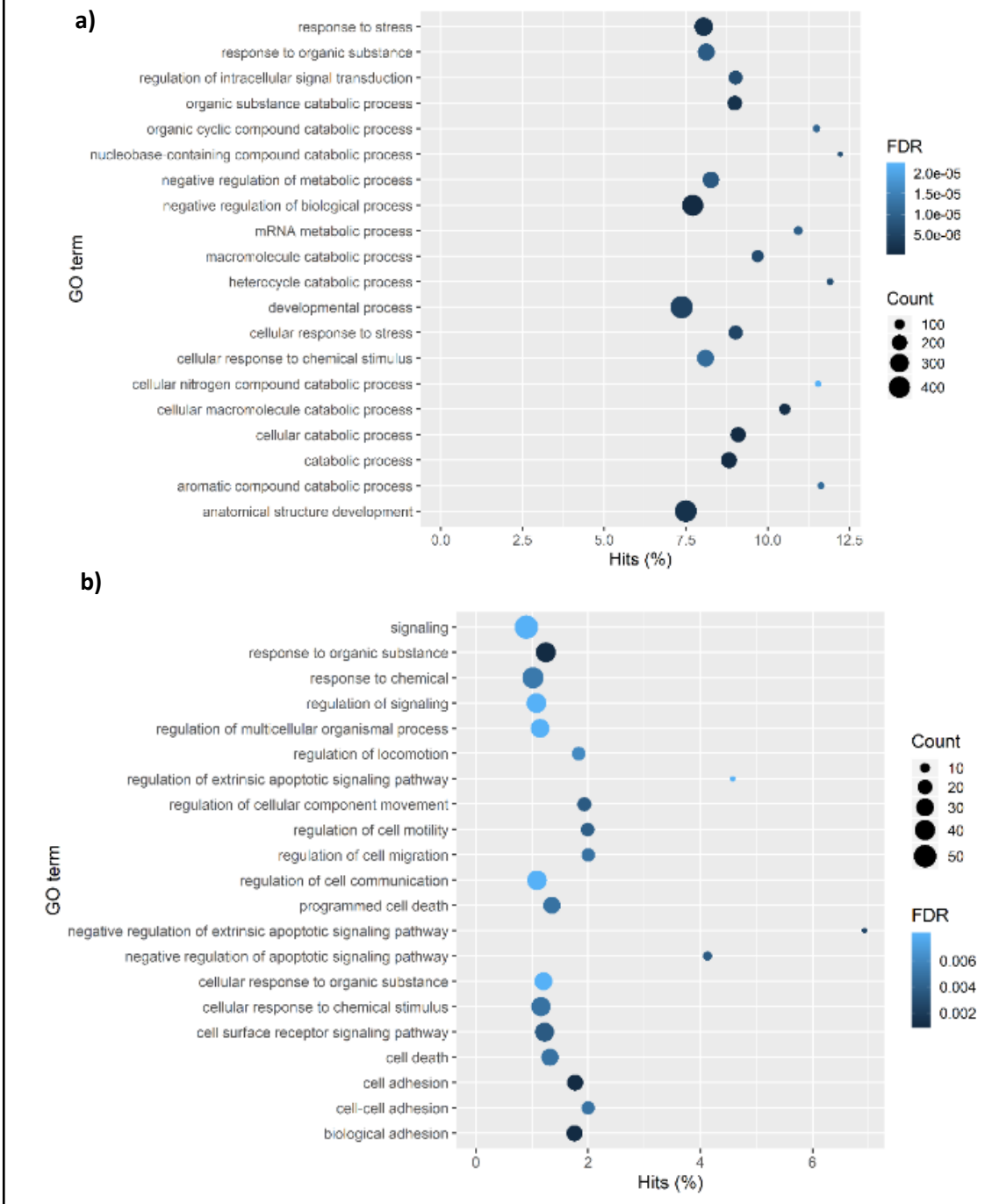


Figure 2. 43- Gene ontology enrichment analysis of DE genes. a) GO analysis on DEG between TRIB1-overexpressed and control samples, irrespective of their time points. **b)** GO analysis of DEG between TRIB1-overexpressed and control samples, at different time points. Dot plot represents top 20 gene ontology annotations; x-axis – the number percentage of differentially expressed genes present in particular ontology function, y-axis – GO annotations; dot size represents the number of differentially expressed genes present in corresponding GO function, and colour represents the FDR score.

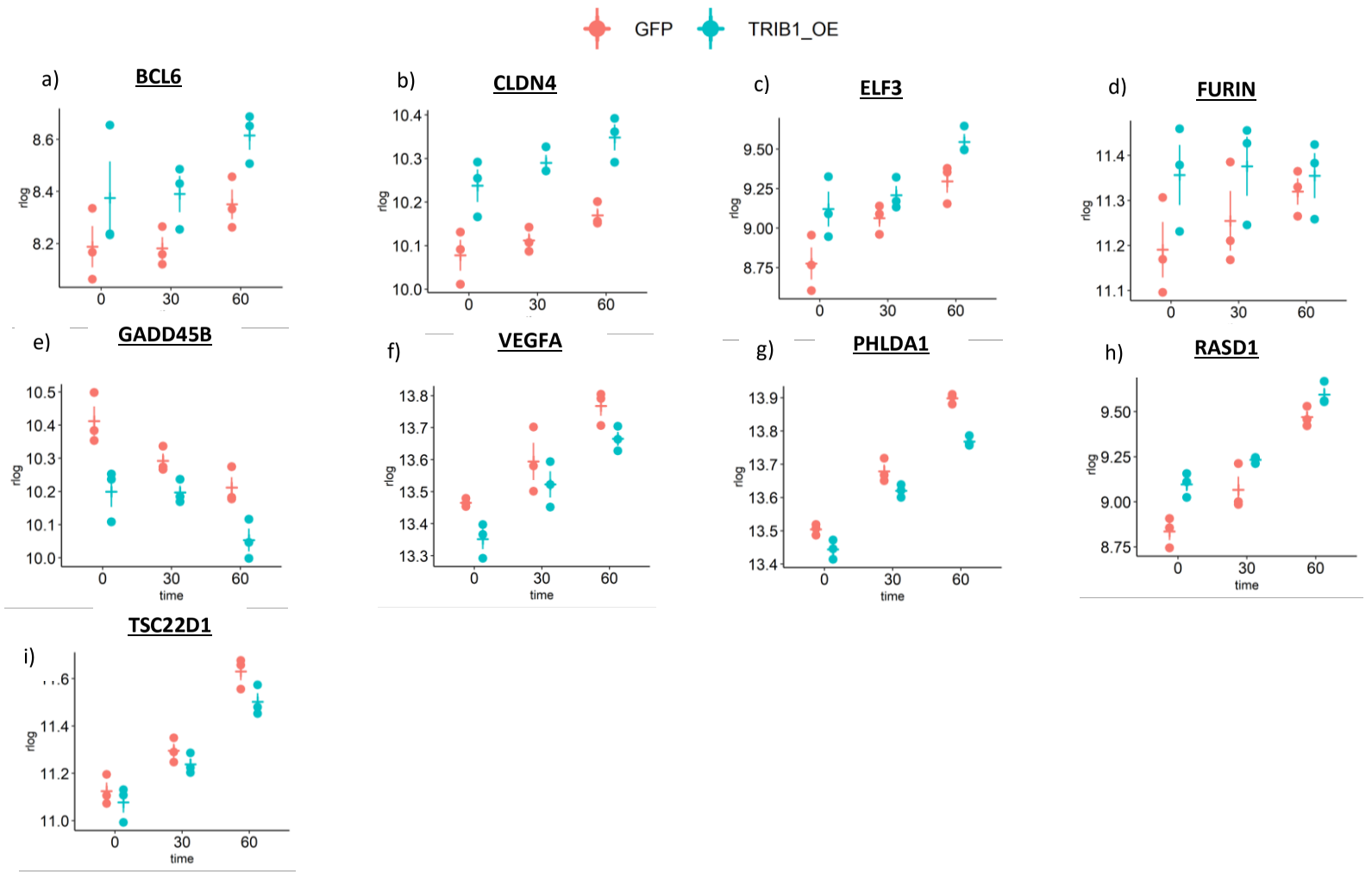


Figure 2. 45- Strip chart showing the change in expression levels of genes co-expressed in prostate cancer with TRIB1 and differentially expressed between control and TRIB1-overexpressing samples, at different time points; *each dot represents a samples and bars represents the mean of the expression values, x-axis – time-points, y-axis – rlog normalized expression value.*

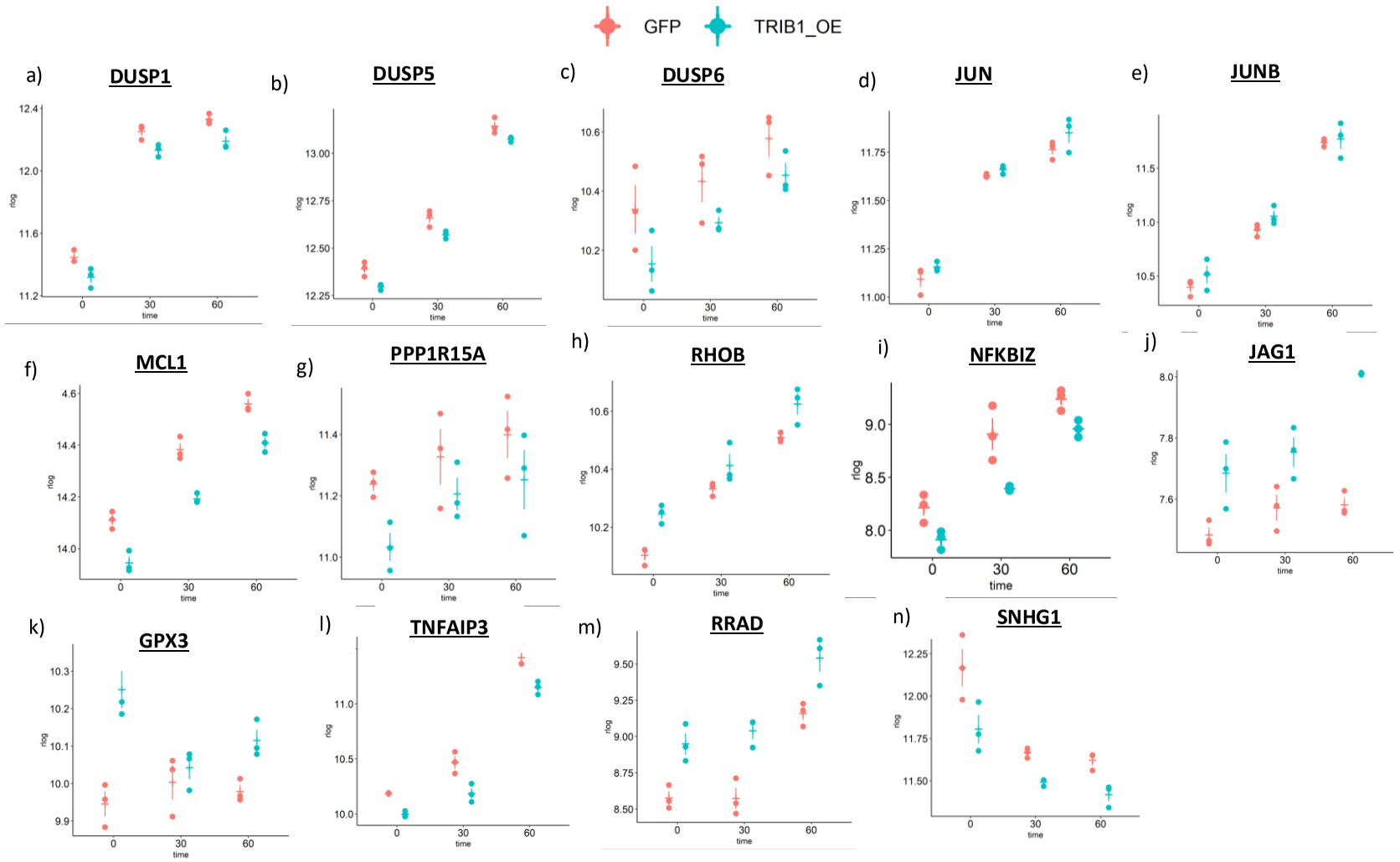


Figure 2. 47- Strip chart showing the change in expression levels of genes co-expressed with TRIB1 in all cancers, and differentially expressed between control and TRIB1-overexpressing samples, at different time points; each dot represents a samples and bars represents mean of the expression values, x-axis – time-points, y-axis: rlog normalized expression values.

2.4 Discussion

In this chapter, we reported that TRIB1 co-expresses with genes with a variety of different functions in macrophages**.** In 21 of 28 cancer, Trib1 co-expresses with a coherent set of genes involved with the immediate early response. Thus we suggest that TRIB1 could be regarded as a that it may in fact be an IER gene. This hypothesis was supported through analyses and results showing that TRIB1 co-expresses, interacts at protein level with and may regulate IER genes EGR1 and FOS; TRIB1 has an IER expression pattern in HEK293T cells, and that TRIB1's overexpression in DU145 cells leads to differential expression of many genes, some of which were identified as early response genes also co-expressing with TRIB1 in prostate cancer.

Co-expression analysis of control macrophages and cells infected with either *Salmonella typhimurium* or *Listeria monocytogenes* showed that TRIB1 co-expressed with different sets of genes between conditions. This could be due to the variations in expression of TRIB1, as TRIB1 was found to be lowly expressed in *Salmonella typhimurium*, compared to control and *Listeria monocytogenes*-infected samples. Considering Trib1 has been previously shown to be involved in macrophage differentiation into an anti-inflammatory M2 phenotype, we investigated the co-expression network of TRIB1 in control MDMs and samples infected with *Listeria monocytogenes*, as infection with this Gram positive bacteria has been found to stimulate an M2-like phenotype in macrophage (Mège, Mehraj, and Capo 2011). Thus, we found TRIB1 to co-express with transmembrane protein TMEM181. This is an interesting finding because TMEM181 together with TRIB1 may regulate the cell signalling pathways and help the cells activate their adaptive immunity. However, the interaction between TMEM181 and TRIB1 needs to be studied. Our analysis is based on the co-expression model and available information in GeneCards (https://www.genecards.org/) of both the genes.

In addition, TRIB1 was also found to co-express with E3 ligase UNKL, and considering past studies have already demonstrated that Trib1 recruits E3 ligase COP1 for the ubiquitination of CEBPA, an interesting possibility is that TRIB1 also interacts may also recruit or sequester UNKL. Indeed, we hypothesise that Trib1 may potentially act as an E2-conjugating enzyme. In support of this, when we analysed the protein structural similarity between Trib1 and E2 ligase UBE2D3, we found a 76% overlap (**Figure 2.4**). This supported our hypothesis but to further validate this relationship, co-immunoprecipitation studies could be performed.

We also investigated the expression of TRIB1 in 28 different cancer datasets and found that 19 of the 28 genes co-expressing with TRIB1 in 18 cancer datasets were early response genes (**Figure 2.9**). We found a interaction between these IER genes and TRIB1 that was at a higher probability than at random, especially for genes co-expressed in at least 10 cancer types. What is more, most of the

co-expressed genes were enriched in cell-signalling pathways, particularly in MAPK signalling, Toll-like receptor pathway and B-cell receptor signalling (**Figure 2.10**). Moreover IER genes like JUN and FOS are known to form the dimeric transcription factor AP-1 (Van Dam and Castellazzi 2001), which has been linked to cancer and was found to regulate gene expression in response to various stimuli, including cytokines, growth factors, stress, bacteria, or viral infections (Wang *et al.* 2013). Furthermore, DUSP1, FOS and JUN show a direct interaction with EGR1, which is widely considered to play an important role in promoting differentiation of macrophages and primary myeloid precursors (Barbieri *et al.* 2018), and has also been found to activate inflammatory enhancers in mature and developing macrophages (Trizzino *et al.* 2021). This observation supports published role of TRIB1 to regulate AP1 dimer. (Kiss-Toth et al. 2004). This our result illuminates the role of TRIB1 in cell-proliferation, cell signalling by regulating early response genes and also supports our hypothesis that TRIB1 may be involved in early cell fate decisions and responses, and that it could indeed be a member of the IER gene family . This could the reason why TRIB1 is not coexpressed with IER genes in macrophages as the data was collected after 2 hours of *salmonella and Lister sp.* Infection.

Out of the early response genes co-expressing with TRIB1, EGR1 and FOS were found in 19 out of the 28 cancer datasets, they have been previously reported to interact with Trib1 at protein level (Ancuta Jurj et al. 2020) (**Figure 2.11**), and they were also recorded to be positively correlated with TRIB1 in the TCGA Firehose RNAseq dataset. In contrast, both EGR1 and FOS were negatively correlated with TRIB1 in DU145 and PC-3 cell lines, as their expression levels were more elevated in DU145 than in PC-3, while the opposite was observed for TRIB1, which was more abundant in PC-3 than in DU145 cells (**Figure 2.13**). Although cell-lines are considered an important tool in studying the molecular biology, despite of their advantages they do not capture the heterogeneity, tissue complexity due to the homogenous group of the cells and their cultivation in controlled environment that lacks the interaction with other cell-types. Also it has been reported other factors that could influence/shows difference of gene expression between cell-lines and tissue is regulation of TF (Pan et al. 2009).

In order to investigate the relationship between these three genes, TRIB1 was overexpressed in DU145 cells and knocked down in PC-3 and HEK293T cells, and the expression of EGR1 and FOS was measured. qPCR analysis results showed an obvious increase in EGR1 expression in TRIB1 k/d conditions in both PC-3 and HEK293T cell lines, while FOS levels were only slightly higher (**Figure 2.14**). This suggested that TRIB1 could inhibit IER genes like EGR1, but this hypothesis was then rejected, because when TRIB1 was overexpressed in DU145 cells, both EGR1 and FOS were again significantly upregulated (**Figure 2.15**), suggesting a more complex regulatory circuit.

In DU145 cells overexpressing TRIB1, EGR1 was significantly increased 30 minutes after early response stimulation with serum, compared to wildtype DU145 cells. It then decreased more rapidly in the overexpressing cells, from 90 minutes onwards. Similarly, FOS was initially elevated in TRIB1 OE condition, but its expression started to significantly plunge earlier, after just 30 minutes of IER stimulation (**Figure 2.16**). These results indicate that Trib1 may have an early effect on early response genes like EGR1 and FOS.

What is more, upon stimulation of IER genes in HEK293T cells by serum, we observed that EGR1 and FOS expression increases from time-point 0 to 60 mins and then starts to gradually decrease. Although the decrease is not as sharp, TRIB1's expression seems to follow a similar pattern, as it goes up for 90 mins post-stimulation and then it is slowly lowered (**Figure 2.12**). This result suggests that TRIB1 could itself be a member of the early response genes family.

Overall, because of the consistent change in EGR1 and FOS levels in response to changes in TRIB1's expression, we cannot completely rule out a regulatory role of TRIB1 over them. However, considering the tests were done in different cell lines, we also cannot conclude on the direction of regulation and the exact mechanism between these three genes.

In addition to its effect on early response genes like EGR1 and FOS, TRIB1's overexpression was also found to lead to the differential expression of several genes at different time points post-stimulation. There were 85 DE genes at time point 0, followed by 30 genes at time point 30 min and 21 DE genes identified at 60 minutes after serum stimulation. However, none of these genes were found to be early response genes, but most were enriched in cell signalling pathways and cell-related biological functions, such as cell migration, mobility and differentiation. In contrast, a larger number of DE genes (1214) was recorded between controls and cells with TRIB1 OE, irrespective of the 3 time points, and they were involved in various metabolic and catabolic processes, as well as responses to stress. These DE genes showed a clear clustering, with more genes being upregulated than downregulated in TRIB1 OE condition compared to DU145 controls (**Figure 2.21**).

Comparing differentially expressed genes detected in TRIB1 OE conditions with genes co-expressed with TRIB1 in the same DU145 cells, we found 18 out of 80 co-expressed genes in prostate cancer (**Figure 2.23**), and 14 out of 28 (**Figure 2.24**) co-expressed in more than 5 cancer datasets excluding prostate cancer were also DE between TRIB1 OE and controls. Most of these genes are early response genes, including DUSP1, DUSP5, DUSP6, JUN, JUNB, MCL1, PPP1R15A, RHOB, TSC22D1, while others play are involved in inflammatory pathways, such as NFKBIZ, TNFAIP3, JAG1, BCL6, VEGFA, ELF3, PHLDA1, FURIN, and few of them have roles in cell proliferation or differentiation, such as RRAD, STC1, MCL1 and GADD45B.

Overall, the above results strongly indicate that Trib1 may regulated by IER genes, and that it is actively involved in regulating cell signaling-related pathways. Also, it suggests that TRIB1 OE may trigger and/or employ distinct transcriptomic behaviours, but further validation in other cell lines could elucidate the exact role of TRIB1, and investigation on the interaction between these differentially expressed genes and TRIB1 could help determine whether there is a direct or indirect effect of TRIB1.

Chapter 3

The role of SNPs altering miRNA binding sites in ASE genes

3.1 Introduction

TRIB1 is known to have a 1.5 kilobase pair (Kbp) partly-conserved 3' UTR, which undergoes miRNA regulation (Lin *et al.* 2014). Previous studies have shown it is a highly unstable transcript and evidence was also found to suggest transcriptional regulation of TRIB1. For example, S. Soubeyrand and colleagues challenged HepG2 cells with a short pulse of a low concentration of oligomycin, after which they observed an increase in the expression of TRIB1 mRNA. However, despite the mRNA upregulation, the levels or stability of Trib1 protein did not increase, which led to the conclusion that TRIB1 is post-transcriptionally regulated (Soubeyrand et al. 2013; Soubeyrand, Martinuk, Naing, et al. 2016). In a separate study, Schwanhäusser and colleagues measured the expression of more than 5000 mRNA and their protein levels in mammalian cells. They found that a random gene's mRNA level correlates with its protein level by 40%, emphasizing that protein expression is controlled by post-transcriptional regulation (Schwanhüusser et al. 2011).

According to the GWA studies, a large number of variations linked to human disease and evolutionary traits were linked to variations in the 3' UTR (Griesemer et al. 2021), and the correlation between genetic variations in these regions and the alteration of miRNAs binding sites has been studied extensively in recent years (Ghanbari et al. 2015). For example Ana Jacinta-Fernandes and collegues reported a variation of approximately 25% in miRNA binding sites between individuals, and part of this variation was also associated with SNPs and correlated with differences in gene expression (Yuan and Weidhaas 2019). What is more, SNPs altering miRNA binding sites can also lead to allelic imbalances, known as allele-specific expression (Jacinta-Fernandes et al. 2020).

All these highlights the importance of studying variations causing alteration in miRNA binding sites. Therefore, this current work aims to provide a more comprehensive view on post-transcriptional regulation of TRIB1 by identifying the effect of variations in 3'UTR on miRNA binding sites and investigating if these variations are linked to allele-specific expression of TRIB1.

3.1.1 Methodological considerations

Over the last decade, there has been a vast improvement in the analysis of SNPs and their effect on gene expression, with many tools being developed for analysing ASE. However, most of the tools available either require known genotype information or trio sequencing, which is expensive and not feasible at large scale. To overcome this difficulty, we developed a pipeline for identifying variants linked to ASE and genes with ASE, without the need of prior genotype information. The pipeline was further applied to analyse the effect of these variants on miRNA binding sites of TRIB1 (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE).

i) Variant discovery and filtering of RNA editing events

The inputs for our pipeline are reference-aligned bam files and raw un-mapped fastq files. The first step includes filtering of the bam files, removing PCR duplicates and splitting the reads based on the mapped exons. Then variants are called using the GATK's Haplotype caller method, which performs de-novo assembly of haplotypes and calls SNPs and indels simultaneously. In other words, when the Haplotype caller finds the discrepancy between reads and the reference due to the variants in the read sequence, it discards the existing mapping information and re-assembles the reads in that region. Moreover, it also corrects the mapping errors made by the original aligner (Poplin *et al.* 2017). Next, RNA-editing sites (RES) are identified from RNA-seq bam files and filtered, using the SNP-free RNA editing IdeNtification Toolkit (SPRINT) (Zhang et al. 2017).

RNA-editing is a modification of RNA molecule, which includes a change in the RNA sequence (i.e. adenosine to inosine and cytosine-to-uracil), with the most common RNA-editing event recorded in humans being adenosine to inosine (Ramaswami *et al.* 2012, Bahn *et al.* 2012). A limitation of traditional RNA-editing identification methods is that they depend on DNA-seq to identify the true variants, or they require using a database of SNPs, such as dbSNPs, to identify RNA-editing sites. However, sequencing both DNA and RNA from the same individual is expensive, and some of the SNPs deposited in dbSNPs have already been reported as RNA-editing sites (Wang *et al.* 2021), which can make identifying new sites difficult. Traditional methods usually follow the below steps to detect RNA editing sites:

1. Aligning RNA-seq files to the respective genomes
2. Identifying SNPs from the aligned files
3. Filtering true SNPs either by DNA-sequencing, or by removing those present in SNPs databases

Unlike the traditional method, SPRINT identifies RE sites without filtering the SNPs, and instead, it distinguishes SNPs from RNA-editing sites by clustering variants duplets (two consecutive SNPs having the same variations: A→G or A→C mutation). This approach was developed taking into account that adenosine deaminases acting on RNA (ADARs) bind to RNA and are responsible for the adenosine to inosine RNA-editing editing event. In addition, RNA editing enzyme apolipoprotein B (ApoB), binds to single-stranded RNA, leading to cytosine-to-uracil editing (Zipeto *et al.* 2015). Although there are more than 150 RNA modifications, both ADARs- and AID/APOBEC-mediated RNA editing events belong to a subset of modifications particular to mRNAs. These are important for RES identification. Moreover, both RES and SNPs events are very unlikely to happen in the same genome location, and it has been observed that adenosine to inosine editing tends to be clustered (Ramaswami *et al.* 2012). On the other hand, other types of SNPs' distribution on the genome is known to be independent (Zhang et al. 2017). Therefore, we incorporated the SPRINT tool into our pipelines, as it identifies the RES events by investigating the distribution of SNP duplets.

Two of the most used approaches for identifying ASE are the binomial method, implemented by Rozowsky and colleagues in the Alleleseq tool (Rozowsky *et al.* 2011), and the analysis of variance (ANOVA) algorithm implemented in the Allim tool (Lu *et al.* 2015). Both these methods, as detailed in Introductory Chapter 1 – sections 1.5.1 a) and b), are based on the difference in read counts between the two alleles. However, depending on the sequencing depth, genes can also have low read counts, which means these methods may ignore genes with true ASE. Therefore, in order to identify ASE, we have implemented two methods based on the Bayesian approach, which share gene information across samples and improve the average on gene-related inferences: first, the QuASAR method determines ASE for each SNP (Harvey *et al.* 2015), followed by the MBASED algorithm, which combines the ASE of each SNP present and identifies ASE at gene level (Mayba *et al.* 2014).

As mentioned in the Introductory Chapter 1 – section 1.5.1, ASE is the imbalance of allelic expression between the diploid copies at the same locus (Salavati *et al.* 2019). The ASE study's biggest challenge in a diploid organism is mapping bias, i.e. reads are more likely to map to the reference allele/haplotype, and reads covering heterozygous loci tend to be erroneously mapped (Hodgkinson *et al.* 2016), which may lead to false-positive results for ASE discovery (Salavati *et al.* 2019). To overcome this challenge, we constructed the maternal and paternal genomes based on the haplotype phasing information. First, we reconstructed the haplotype information for each SNP by using the `Readbackphasing` tool from GATK, which is one of the few tools that support vcf files as an input and output. It uses the physical read information, as reads are considered with a Bayesian framework and the algorithm then constructs the haplotype for each SNP based on its probability (number of reads observed for particular SNP) (Depristo *et al.* 2011) (**Figure 3.1**).

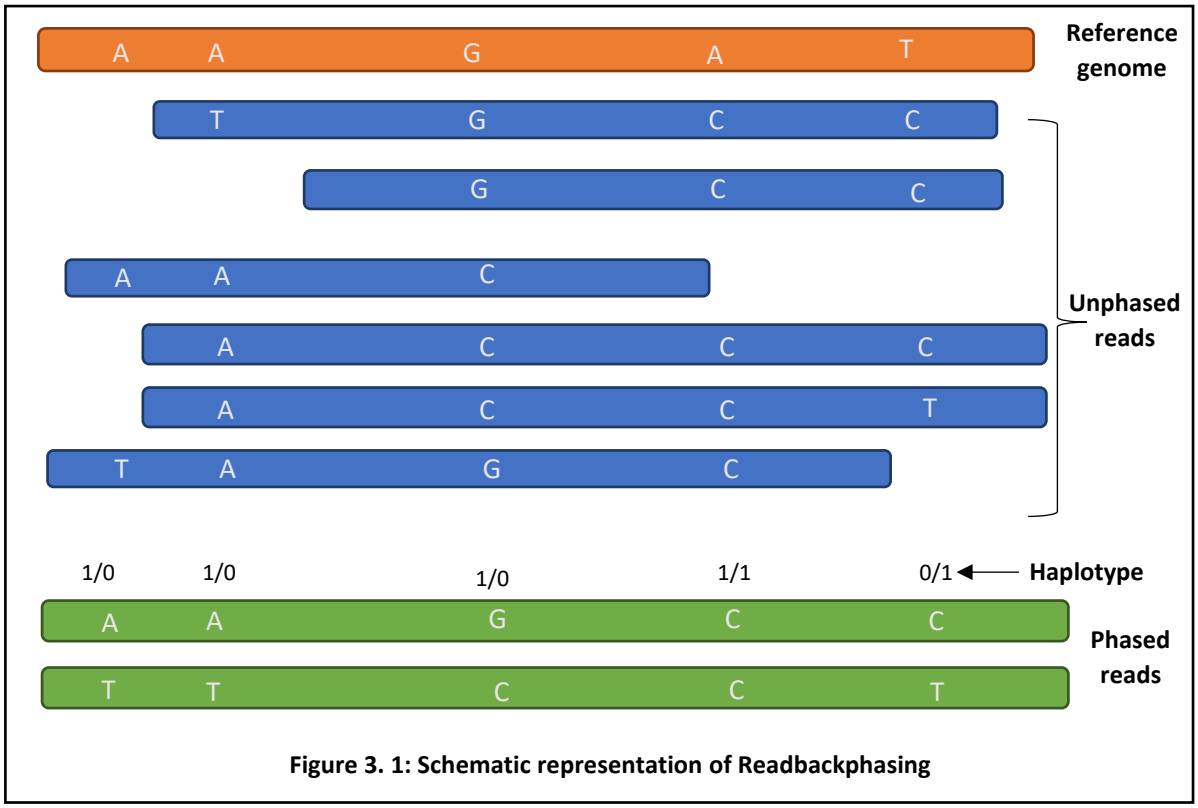


Figure 3. 1: Schematic representation of Readbackphasing

ii) Predicting effects of variation on miRNA binding sites

In order to predict miRNAs targeting transcript variants with SNPs in their 3'UTRs, we used miRanda, which first takes into consideration the sequence complementarity between the mature miRNA and the target site. Next, it predicts the binding energy of the miRNA–target duplex, and it then (optionally) looks for conserved 6-8 seed sequences of miRNA binding sites in 3'UTR of the homologous gene target. (Betel et al. 2010)

3.2 Materials and methods

3.2.1 Datasets

As we aimed to identify sequence variants in the 3' UTRs, their effect on miRNA binding sites and their correlation to allelic imbalance, and considering that commonly available exome sequencing data does not include 3' UTRs, we turned to RNA-seq data for our analyses.

RNA-Seq (Human) dataset (accession number GSE81046) was obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) and the associated fastq files were downloaded. The dataset included 169 samples prepared from monocyte-derived macrophages (MDMs) of control individuals and 96 samples challenged with *Salmonella typhimurium* infection (Nédélec *et al.* 2016). The reason we selected a *Salmonella typhimurium*-related dataset was because this type of infection is usually associated with an M1 pro-inflammatory phenotype in macrophages (Lathrop et al. 2018), which is also triggered by LPS-IFNγ stimulation that we used in our *in vitro* experiments.

To check the ASE pipeline's efficiency and accuracy, we downloaded NA12878 cell line (Only cell line with known genotype information as per my knowledge) RNAseq from the GEO database (accession number GSE30400).

3.2.2 Quality check and mapping

`FastQC` (Andrews S, 2010) was used to check the quality of raw reads. Adapter sequences were trimmed using the `Trimmomatic` tool (Bolger, Lohse, and Usadel 2014) using the following parameters:

```
LEADING:3 \
TRAILING:3 \
SLIDINGWINDOW:4:15 \
MINLEN:36 \
Mismatch: 1 \
Phread:30
```

Raw files were then mapped to the human genome (hg38) using STAR with default parameters (Dobin *et al.* 2013).

3.2.3 ASE pipeline

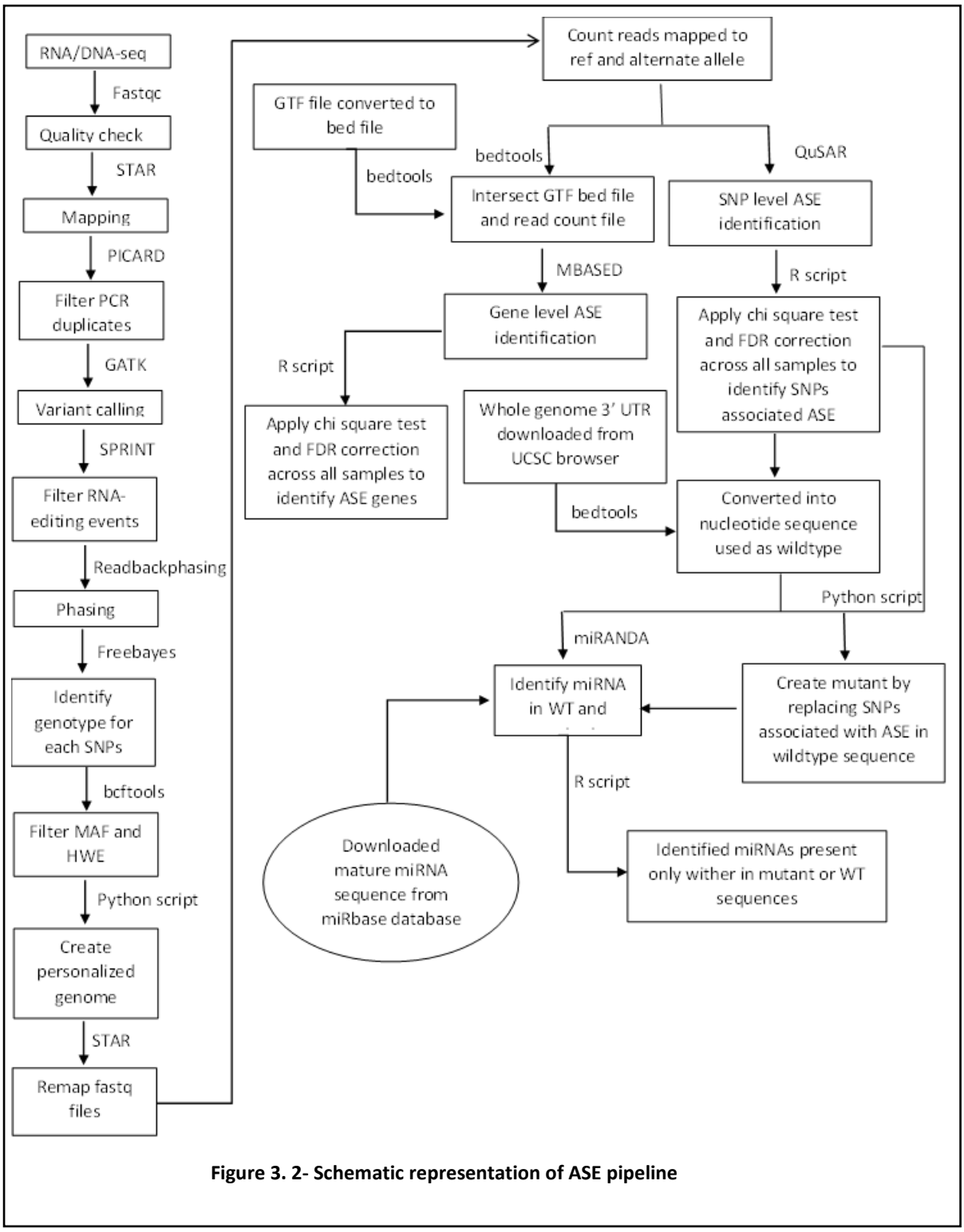


Figure 3. 2- Schematic representation of ASE pipeline

i) Reprocessing aligned reads for variant detection

`Picard` (http://broadinstitute.github.io/picard/) was used to mark duplicate reads generated by PCR amplification during library preparation. Next, reads mapping to different exons were split, using `SplitNCigarReads` from the GATK toolkit (**Figure 3.3**). Then, the mapping quality score was recalibrated with `ReassignOneMappingQuality` parameter, as `STAR` alignment assigns a good alignment score (MAPQ) of 255, a value that GATK does not recognize.

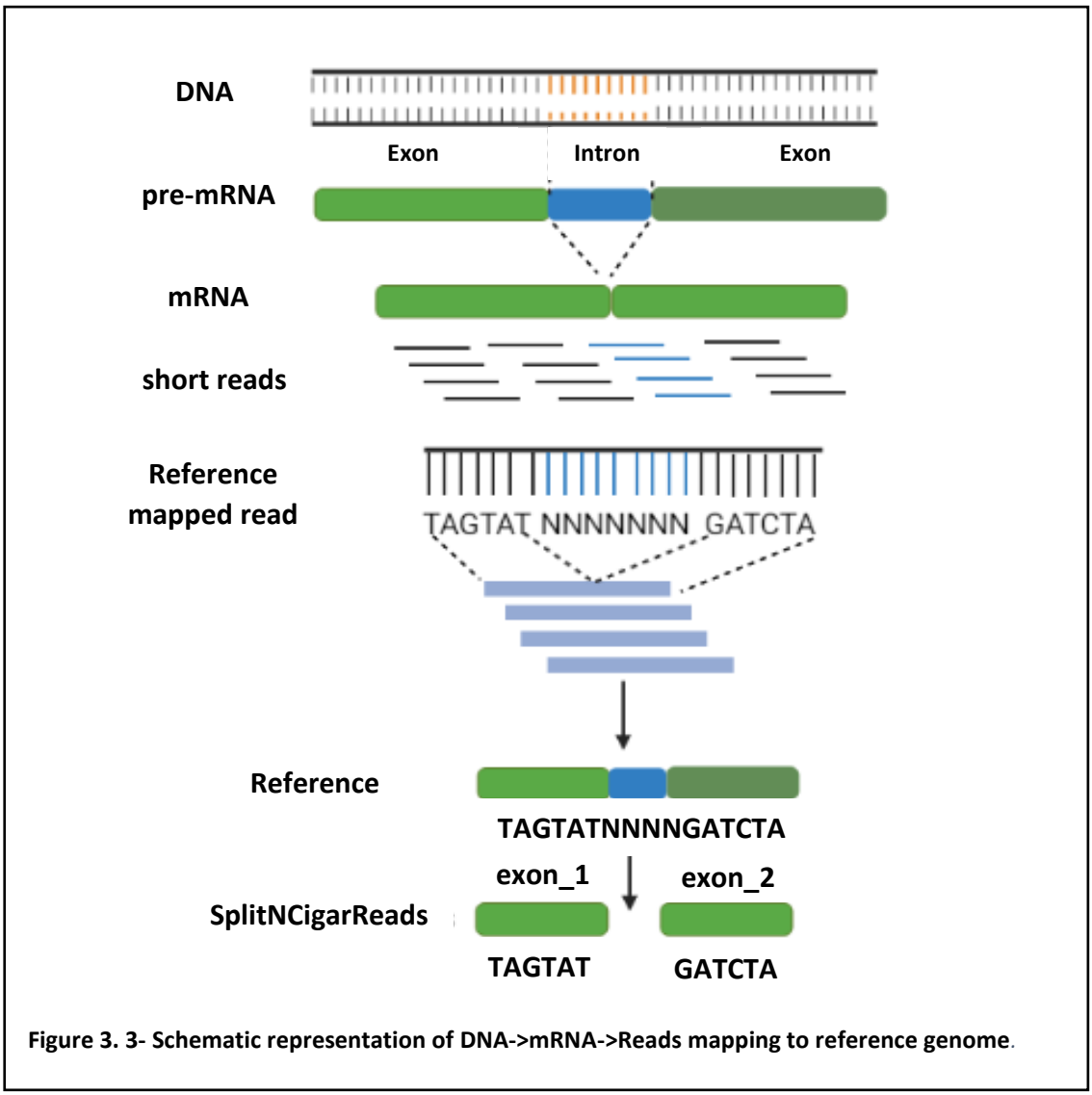


Figure 3. 3- Schematic representation of DNA->mRNA->Reads mapping to reference genome.

ii) Variant calling

The final aligned file generated from the previous step was then parsed to call variants, using the GATK `HaplotypeCaller` (Poplin *et al.* 2017), with the parameters below, which considers the minimum Phred-scaled confidence equal or greater of 20:

```
-dontUseSoftClippedBase\
  stand call conf = 20
```

For diploid organisms (e.g. human), GATK assigns 0 for reference allele and 1 for the alternate allele (non-reference) and the resulting genotypes can be heterozygous (0/1 or 1/0 – carrying both reference and alternate allele) or homozygous alternate (1/1).

iii) Filtering RNA-editing sites

In order to identify RNA-editing sites, we use the `SPRINT` method, with the following steps:

1. Different reference genomes were created by replacing T → C and A → G, for identifying A → G and T → C RNA editing sites from RNA-seq reads
2. A genome index was separately generated for each reference genome, using Burrows-Wheeler Aligner (Li and Durbin 2009)
3. RNA-seq reads were mapped to all the genomes using BWA, and output as .sam files
4. Aligned sam files were converted to bam files with *samtools* (v1.2) (Li and Durbin 2009)
5. Duplicates were marked using Picard-tools (http://broadinstitute.github.io/picard/)
6. Reads with mapping quality > 20 were considered as mapped reads
7. SNVs were called on mapped reads, and RNA editing sites were identified based on the SNV duplets clustering method
8. Unmapped reads were masked by replacing A → G and remapped to the masked genome (Wang *et al.* 2021)
9. The final output was a text file containing the RNA-editing sites, as shown in **Table 3.1**

Table 3 1- Output produced by SPRINT toolkit. . Column: 1 – chromsome name, 2 & 3 – start and end of the genomic location for the RE sites, 4 – type of RNA-edit (A → G ot T → C), 5 - number of reads at the site, 6 – strand

Chrom	Start	End	Type	Supporting reads	Strand
chr1	1608103	1608104	AG	10	+
chr1	1608104	1608105	AG	22	+
chr1	1608106	1608107	AG	23	+
chr1	1608108	1608109	AG	1	+

SPRINT's output was further passed to a custom R script (Convert_bed_rnaediting.R, https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/Convert_bed_rnaediting.R) to convert it into a bed format. RNA-editing sites were

```
bedtools subtract -a vcf file \
                  -b outfile_from_STRINT \
                  -header -s > vcffile
```

then filtered from the vcf file generated in the previous step using *bedtools* (Quinlan and Hall 2010):

Next, indels and multiallelic SNPs were filtered using *bcftools* (Narasimhan *et al.* 2016):

```
bcftools view --max-alleles 2 \
              --exclude-types indels \
              input.vcf.gz > filtered.vcf
```

The haplotype was then reconstructed from the filtered .vcf files, as explained below.

iv) Phasing

Haplotype information for each SNP was reconstructed with the `Readbackphasing` tool from GATK, using the following command:

```
java -jar GenomeAnalysisTK.jar -T ReadBackedPhasing \
        -R reference_file \
        -I Aligned_file \
        --variant variant_calling file \
        -L variant_calling file \
        -o output.vcf \
        --phaseQualityThresh 20.0 \
        --allow_potentially_misencoded_quality_scores \
        -allowPotentiallyMisencodedQuals
```

v) Constructing personalized genomes

A custom Python script (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/generate_split_genomes.py) was then used to generate two genome sequences for each individual. The inputs required for the `Splitgenome.py` script are the reference_genome.fasta and the phased .vcf file. Based on the phasing information from the .vcf file, `Splitgenome.py` then generates the personalized genome of parents for each sample.

vi) Remapping to the personalized genome

The original .fastq files were re-mapped to the personalized genome using `STAR`, with stringent parameters, which allowed only one mismatch per read:

```
--outFilterMultimapNmax 1 \
-- outFilterMismatchNoverLmax 0.05
```

Next, an alignment file was generated that retained only the reads where the best mapping results from either one of the personalized genomes or the other. This was done through another Python script,`filter_bam.py` (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/filter_bam.py).

vii) Identifying the genotype for each genomic position and minor allelic frequency (MAF)

Variant calling was performed with the GATK's `HaplotypeCaller` and then `Freebayes` (Garrison and Marth 2012) was used to identify the genotype for each SNP in each sample (as GATK doesn't return genotypes for reference homozygous positions):

```
Freebaye  -f standard reference genome \
          -@ Filtered_varaiant_calling file \
            Aligned_bam_file >  output_file.vcf
```

Next, multiple .vcf files were merged using `bcftools` with command :

```
bcftools merge (all_vcf_files) > merged.vcf
```

Finally, minor allelic frequency was calculated for each sample set using the `bcftools` plugin `fill-tags`, with the following parameters:

```
bcftools +fill-tags infile.vcf -o output.vcf -- \
-t AF, HWE, MAF
```

where **AF** is Allelic frequency, **HWE** is p-value from a chi-squared test against Hardy-Weinberg equilibrium, **MAF** is minor allelic frequency, and the formula for calculating MAF by `bcftools` is explained in **Table 3.2**.

```
vcftools --vcf %(infile)s --hwe 0.01 --recode --out %(output)s
```

Table 3 2- Theoretical example of MAF calculation. 0/0 = homozygous reference (counted as 0); 0/1 = heterozygous (counted as 1); 1/1 = homozygous alternative (counted as 2); Every SNP has 2 copies of each allele and we have a total of 3 samples for SNP1, therefore MAF calculation is: (0+1+2)÷6 = 0.5

SNPs	Sample1	Sample2	Sample3
SNP1	0/0	0/1	1/1
SNP2	1/1	0/1	0/1

viii) Input for ASE analysis

The number of reads mapping to the reference allele, alternate allele, or any other nucleotide positions were retrieved using a customized Python script (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/base_count.py). The Python script required an input filter vcf file (section

3.2.3 iii) Filtering RNA-editing sites), personalized genome fasta file, filtered bam file (section 3.2.3 v) Constructing personalised genome), and vcf file with a minor allelic frequency (section 3.2.3 ii) Variant calling at each genomic position and nominal allelic frequency) and the output generated as a bed file, with an example shown in **Table 3.3**.

Table 3 3- Example of input bed file for allelic-specific expression tool.*;* column 1 – chromsome number, columns 2 and 3 – SNP genomic position, columns 3 and 4 – nucleotide at reference and alternate positions, column 6 – dbSNP id, column 7 - Minor allelic frequency, columns 8,9 and 10 – number reads mapping to reference, alternate and or any other nucleotide.

Chr1	16494	16495	G	C	rs3210724	0.5	18	52	0
Chr1	184990	184991	G	A	rs1219494595	0.16	409	76	0
Chr1	185794	185795	G	T	rs1271744271	0.5	56	338	0

ix) ASE at SNP and gene levels

QuASAR R package was employed for calculating ASE at the SNP level using R script `Quasar_output.R`, (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/Quasar_output.R) , which took as input the bed file generated in the previous step (**Table 3.3**) and then generated an output file for every sample. To determine SNPs with a reproducible effect, P-values for the same SNP across only those samples with that were heterozygotes for the SNP were combined using Fisher's method, and multiple testing correction for the combined P-value was then calculated with Benjamin-Hochberg procedure.

In order to identify ASE at gene level, we used MBASED, with input files generated as follows:

a. Gene transfer format (GTF) files hold genomic information in tab-delimited format, such as genes, gene IDs, transcript IDs, protein names, strand and genomic features (exon, intron, START and STOP codons) etc. GTF files were converted into a bed format, using `gtf2bed` from BEDOPS tools (Neph *et al.* 2012), and the first four columns containing the gene names, coordinates (start and end positions) and chromosome names were extracted, generating a bed_gtf_file.bed as an output:

```
gtf2bed < GTF_file | cut -f1-4 > outfile
```

b. The bed_gtf_file.bed file generated from the previous step and the bed file generated in the previous section (bed_input_ASE used as an input for allele-specific expression) were merged using `bedtools` as follows:

```
bedtools intersect -a bed_input_ASE \
-b bed_gtf_file \
-w ab > MBASED_input_format
```

c. The R script `gene_levelMBASED.R` (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/gene_levelMBASED.R) was then used to call the MBASED function on each sample separately

To identify the effect of ASE on genes, we combined the p-values for the same gene across all samples using Fisher's method, and multiple testing correction for the combined P-value was calculated using Benjamin-Hochberg.

3.2.4 Validation of the ASE pipeline

We tested the performance of our ASE pipeline against the known genotype information for NA12878 RNA-seq data set (Rozowsky et al. 2011) by using the Alleleseq script to generate genotype as mentioned in their paper, as it is one of the few tools with a complete workflow for identifying ASE. Hence, the RNAseq data was processed according to the Alleleseq approach and it was mapped to the human genome (hg38). Each sample's parental genome was generated using the *vcf2diploid* function of Alleleseq, requiring the phased genotypic information from parents, which was downloaded from 1000 genome project databases.

We then mapped the NA12878 RNA-seq samples to the personalized genomes generated, and the aligned files were processed *via* a customized Python script (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/base_count.py) to count the reads mapped to the reference and alternate alleles of the SNPs. These were further used as inputs for QuASAR to identify the ASE at SNP level and the gene information was added to the SNP file generated. These files were further passed to MBASED to identify gene-level ASE.

3.2.5 Identification of altered miRNA binding sites

We downloaded mature miRNA sequences from the miRbase database (Kozomara, Birgaoanu, and Griffiths-Jones 2019), and a bed file with whole-genome 3' UTRs locations was downloaded from the UCSC table browser (Karolchik *et al.* 2004). Next, RNA editing sites for each individual were filtered to keep only true variants in 3' UTRs, using *bedtools subtract*:

```
bedtools subtract -a vcf_file \
-b outfile_from_STRINT > vcffile
```

The bed file obtained from the UCSC browser was also used to retrieve the corresponding fasta sequences using *bedtools getfasta:*

```
Bedtoools getfasta -fi reference genome \
-bed bedfile > wildtype sequence
```

A customPython script (https://github.com/srmeetd/Pipeline_Allele-specific-expression-ASE/blob/main/generate_split_genomes_utr.py) was then used to mutate the wildtype fasta sequences generated in the previous step, taking the filtered vcf file for each individual mentioned above as inputs. This script mutates the wildtype fasta sequences by replacing reference nucleotides with the alternate at the particular genomic positions present in the vcf file.

Using both the reference and the mutated fasta files, putative miRNA binding sites were then predicted with the miRNA binding site prediction algorithm miRanda (Betel *et al.* 2010), using the parameters *en -20 -strict*, where *en* is energy score and *strict* is alignment with only 7-nucleotides seed sequences. For achieving this, miRanda queries both wildtype 3' UTR and mutant 3' UTR fasta sequences and scans them against miRNAs downloaded from the miRbase database. Next, we developed an R script (*compare miRNAsBS.R,* https://github.com/srmeetd/miRNA_bindingsites) to compare the miRNA binding sites between wildtype and mutant, which looks at wild-type and mutant miRNAs binding site coordinates, miRNA names, and transcripts generated by the miRanda algorithm. It then compared the miRNAs binding between wildtype and mutant for each transcript, and reported the novel miRNA binding sites created and/or the existing miRNA binding sites destroyed.

3.2.6 Gene ontology and Linkage disequlibrium analysis

First, we extracted 3'UTR sequences from hg38 GTF file, calculated their length and then removed length biased 3'UTRs. Gene ontology enrichment analysis was performed using GOseq (Young et al. 2010) R package. Next, FDR was calculated on the 'over_represented_pvalue' column output obtained from GOSeq. The reason for using the 'over_represented' column is that it represents more DE genes in the present categories than expected at any given size of category and length distribution; hence, these are considered enriched DE genes in that category (Young et al. 2010). Gene ontology passing the threshold of FDR < 0.05 were then selected for further analysis. We then calculated SNPs in LD with an R2 ≥ 0.6 and 200 kb distance threshold with the SNPs linked to allelic imbalance using the LDlink webtool (Machiela and Chanock 2015).

3.3 RESULTS

3.3.1 Variation in the 3' UTR of TRIB1

3.3.1.1 Sequence variants were identified in TRIB1 3'UTR

Quality control analysis on RNAseq from 169 samples of unstimulated monocyte-derived macrophages showed raw reads were of good quality, which were directly mapped to the human genome, without further processing. On average, 94% of reads were uniquely mapped. After performing deduplication, variant calling and filtering indels, multi-allelic variants and RNA-editing events, 58,661 genotype variants were identified, with 24,201 heterozygous (carrying both reference and alternate alleles), 11,764 homozygous reference and 22,696 homozygous alternate. 60 of these variants were located in the TRIB1 3'UTR and at least one variant was detected in 156 patients.

Furthermore, in order to identify the half maximal number TRIB1 variants detected in most individuals, we used the below formula:

$$(min.number\ of\ variants\ in\ one\ sample + max.number\ of\ variants\ in\ one\ sample)\ /\ 2$$

Overall, we observed that the highest number of variants detected in any given sample was 13, and the minimum was 1, leading to an half-maximal value of 7 variants of TRIB1 (**Figure 3.4**, **Table 3.4**). This value was then used as a threshold to count the number of individuals with more than the half maximal number of variants. As we can observe in **Table 3.4**, there are significantly less samples with more than 7 variants, which can help identify sample-specific SNPs. More than that, higher number of SNPs is also more likely to cause alterations in miRNA binding sites, which can ultimately lead to a change in gene expression levels.

Figure 3.5 shows the expression of TRIB1 in samples with more than 7 variants, sample with more than one and less than seven variants, samples with only one variant and samples with no variants. Our results indeed suggest that the expression of TRIB1 could increase with the number of variants, which could be due to the alteration in miRNA binding sites in TRIB1 3'UTR, and it motivated further down investigation of the relationship between SNPs and miRNA binding sites. Alternatively, it may also mean more variants can be discovered in samples with higher expression of TRIB1.

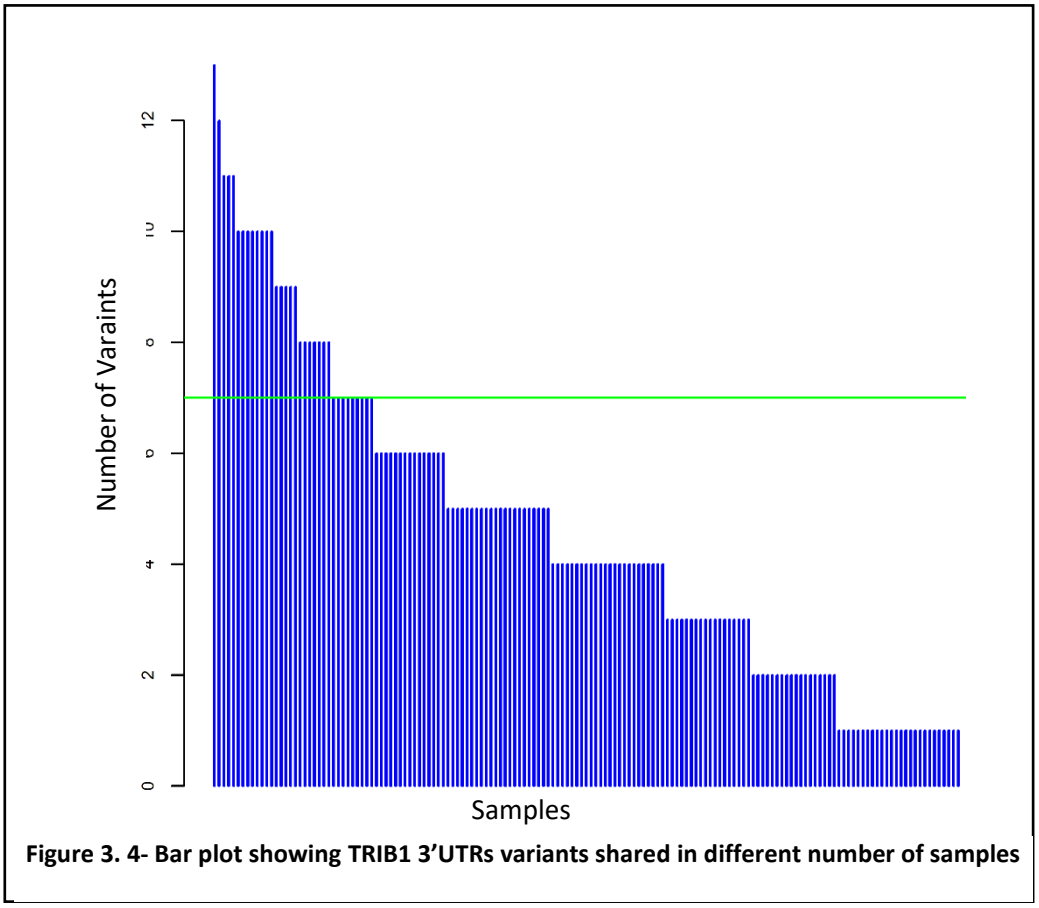


Figure 3. 4- Bar plot showing TRIB1 3'UTRs variants shared in different number of samples

3.3.1.2 Variants in 3'UTR of TRIB1 alter miRNA binding sites

After scanning miRNA sequences from miRbase database against both wild-type and mutant TRIB1 3'UTR sequences (as described in section 3.2.5 Identification of altered miRNA binding sites), we found a total of 1237 predicted miRNAs binding sites present in 3' UTR of TRIB1, out of which 11 variants destroyed 20 miRNA binding sites across 24 samples (**Figure 3.6**), while 18 variants created novel binding sites for 38 miRNAs across 39 samples (**Figure 3.7**).

Table 3 4- Number of TRIB1 3'UTR variants detected in a given set of samples

No. of variants	No. samples
1	26
2	18
3	18
4	24
5	22
6	15
7	9
8	7
9	5
10	8
11	3
12	1
13	1

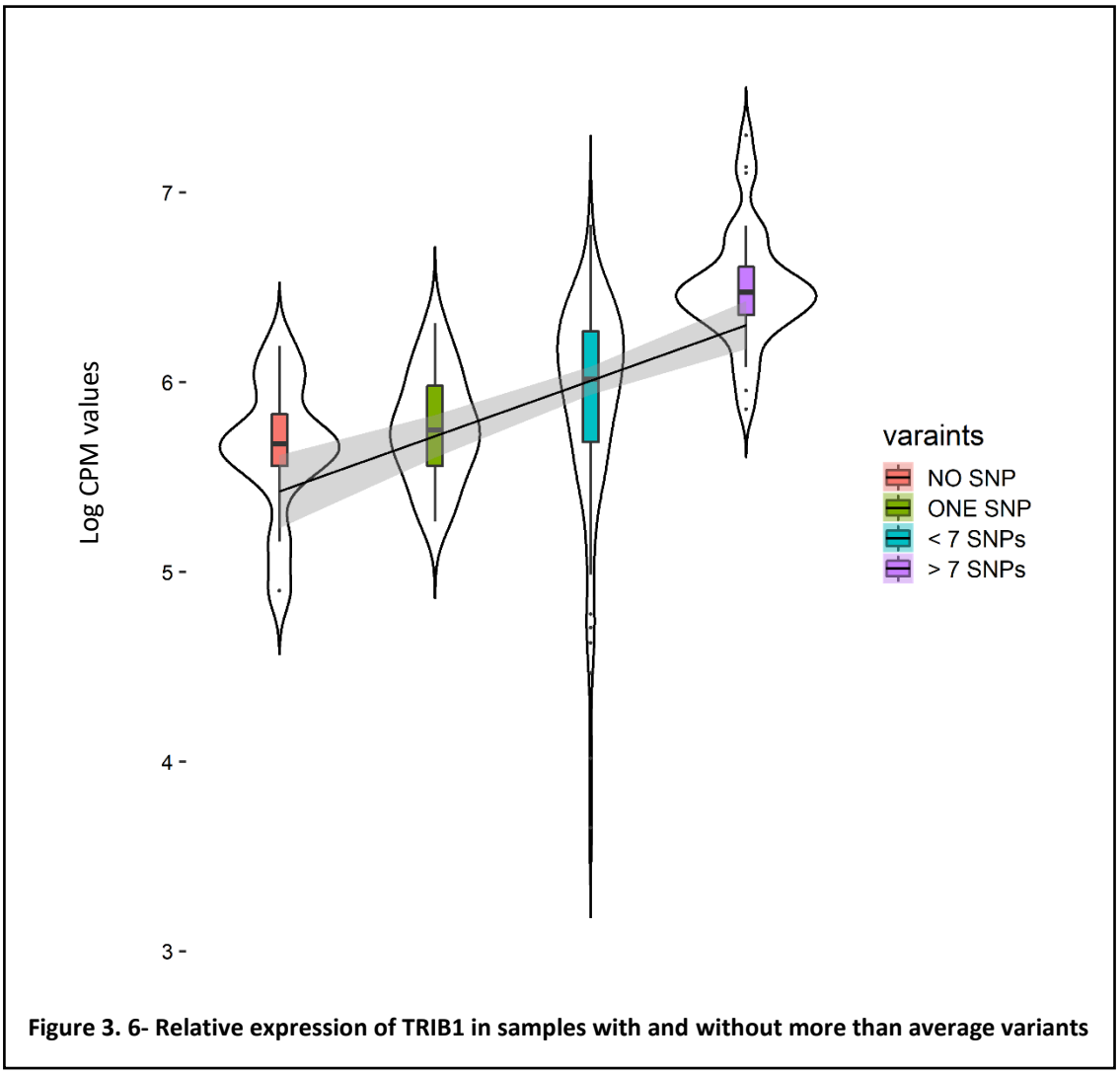


Figure 3. 6- Relative expression of TRIB1 in samples with and without more than average variants

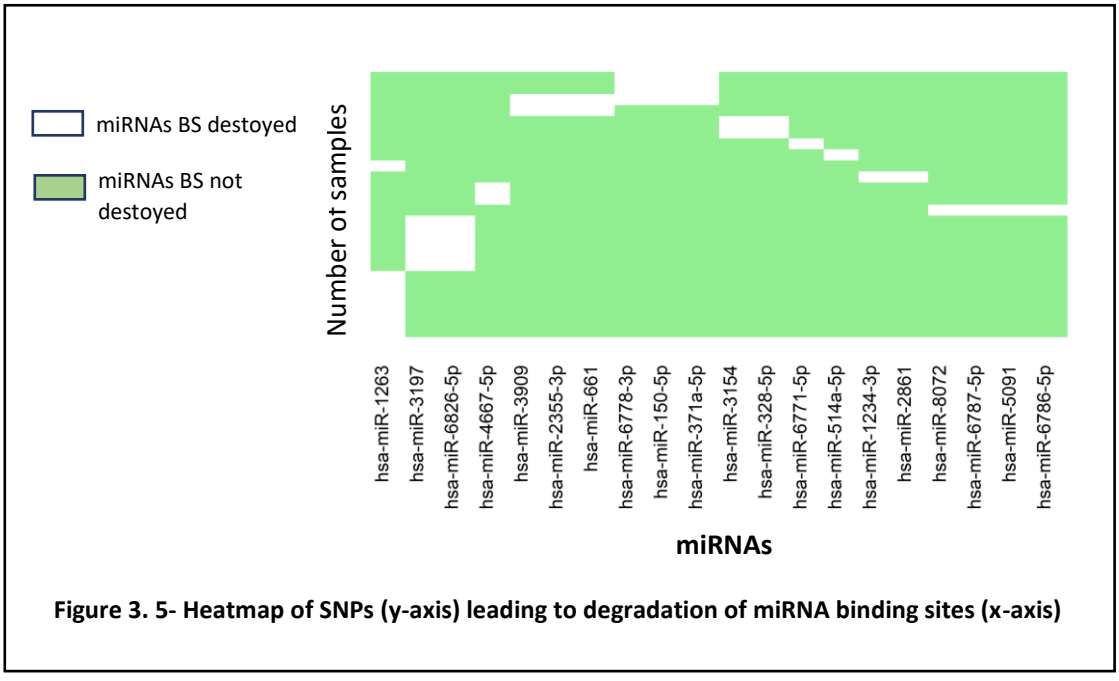


Figure 3. 5- Heatmap of SNPs (y-axis) leading to degradation of miRNA binding sites (x-axis)

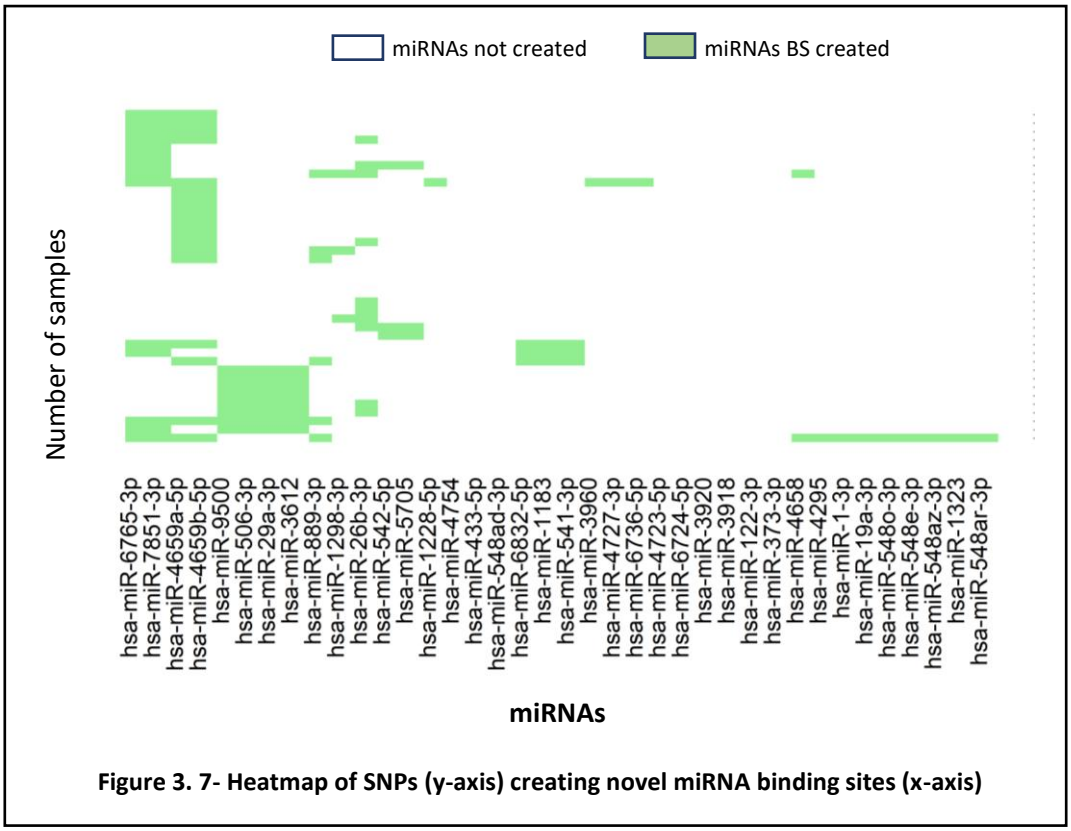


Figure 3. 7- Heatmap of SNPs (y-axis) creating novel miRNA binding sites (x-axis)

Moreover, the expression of miRNAs whose binding sites were altered by SNPs in TRIB1 3' UTR was also investigated in small-RNAseq from unstimulated and LPS+Ifg challenged (M1-like) macrophages (discussed in Chapter 4). Nine (9) miRNAs – 3 with destroyed binding sites and 6 with novel binding sites created, were detected in the macrophage small-RNA sequencing dataset, which suggests that SNPs in 3'UTR are more likely to create new putative miRNA binding sites than to destroy existing ones. However, very few miRNAs were expressed in macrophages and only one of them was found to be differentially expressed between unstimulated and M1-like macrophages, miR-29a-3p.

3.3.1.3 rs62521034 SNP creates a novel binding site for miR-29a-3p

To validate our computational method's accuracy for detecting the effect of SNPs on miRNA binding sites, one of our collaborators, Dr. Chiara Niespolo in the IICD department of the University of Sheffield performed the relevant, corresponding *in-vitro* experiments. She used a TRIB1 mutant containing a T to C base change in the 3'UTR sequence (rs62521034), which was predicted to create a binding site for hsa-miR-29a-3p miRNA, which does not target WT TRIB1. This variant sequence was then expressed in HEK293T cells, together with a miR-29a-3p mimic and their expression was tested via a dual-luciferase reporter assay. The results showed a decrease in luciferase intensity for the

mutant 3'UTR compared to the WT, thus confirming the creation of a novel binding site for this particular miRNA (**Figure 3.8**).

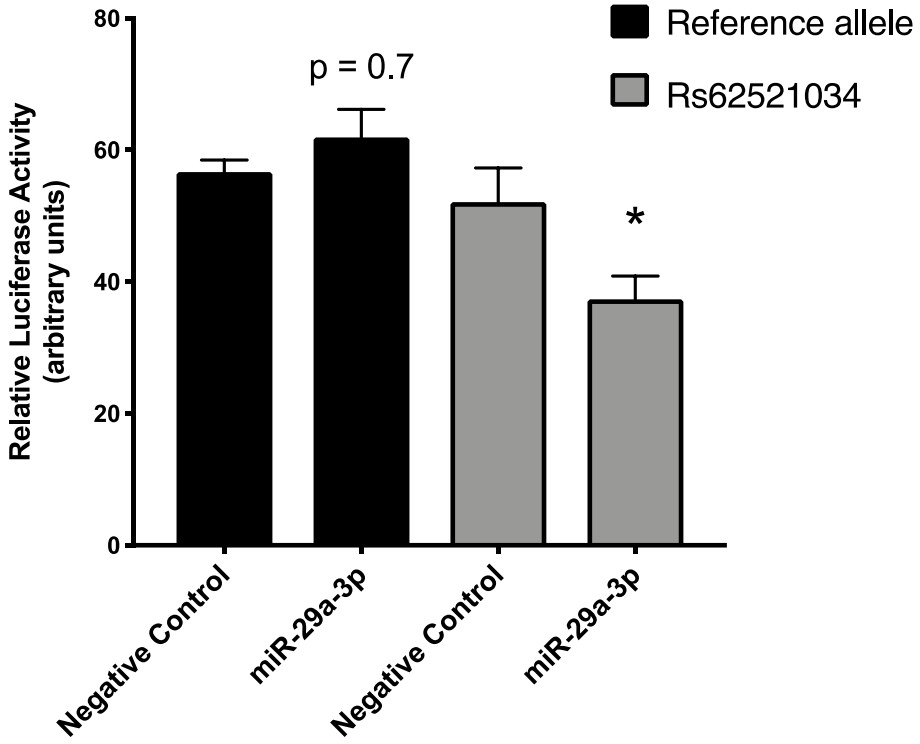


Figure 3. 8- Relative luciferase activity in HEK293T cells transfected with the WT and rs62521034 mutant TRIB1 3'UTR-reporter. Cells were overexpressed for 24h with miR-29a-3p and a miRNA negative control (scrambled miRNA). (mean ± SEM, * $p<0.05$, One Way ANOVA, n = 5)

3.3.2 The effect of miRNA binding site variation on Allele-Specific Expression (ASE).

In order to further investigate if SNPs in 3' UTR of TRIB1 are linked to ASE, We/I developed a CGAT pipeline. Our pipeline identified ASE events at gene and SNP levels using RNA-seq and DNA-seq, even when prior phasing information is unavailable. Moreover, our pipeline also identifies SNPs linked to ASE responsible for altering miRNA binding sites.

3.3.2.1 Our ASE pipeline was successfully validated

To test our pipeline's effectiveness, we implemented the vcf2diploid function of Alleleseq, which is only currently published and well-cited method to generate parental genome to identify ASE events is to our knowledge. The limitation of vcf2diploid function is it requires phased genotypic information. Therefore, we used NA12878 RNA-Seq data as it's true phased genotypic information is available on 1000G project datatabase. Each sample of NA12878 RNA-seq dataset was mapped to the personlozed genome generated using vcf2dipploid (**Table 3.5**). Next, parental genome were generated through ASE pipeline and each sample of NA12878 were mapped. We observed a comparable results between our own ASE pipeline (no prior phasing) and the vcf2diploid (Prior phasing), such as similar mapping percentages, as recorded in **Table 3.5**.

Table 3 5 - RNA-seq mapping percentages obtained from ASE pipeline and vcf2diploid from known genotype (used in Alleleseq)

RNA-seq mapping		
Reference (hg38)	79.65%	
	ASE pipeline	Known genotype
Parent 1	78.64%	76.66%
Parent 2	78.65%	76.65%

The major differences in approach of two pipelines are described in **Table 3.6**, with the two most striking ones being 1) the variant calling step integrated in the ASE pipeline, but not required for Alleleseq, and 2) the DNA phasing information which is not needed for our ASE pipeline, but is essential for Alleleseq

Table 3 6- Major difference between ASE pipeline and the Alleleseq tool

	ASE pipeline	Alleleseq tool
Data (RNA-seq and DNA-seq)	Yes	Yes
Variant calling	Yes	No
DNA Phasing info required	No	Yes
Diploid genome creation	Yes	Yes

Next, we detected 1692 SNPs by QuASAR using prior genotype information (**Figure 3.9a**), out of which 36 were significant, with an FDR < 0.05, thus being linked to ASE (**Figure 3.9b**). What is more, a total of 256 genes showed > 1 heterozygous loci, out of which 180 were associated with ASE, having passed the cutoff of FDR 0.05 set in MBASED, also part of our ASE pipeline. By comparison, using QuASAR in our ASE pipeline, but without prior genotype information, we found 2032 SNPs, out of which 39 were significant, with an FDR < 0.05, thus being linked to ASE. Similarly, by implementing MBASED, 302 genes were found to have > 1 heterozygous loci, but only 139 passed the cutoff FDR < 0.05, being linked to ASE. (**Figure 3.9**)

The overlapped results at both SNP and gene levels recorded through our ASE pipeline with and without prior genotype information are shown in **Figure 3.9** and Supplementary Table 1 (https://github.com/srmeetd/ASE_supplemantry_DATA/blob/main/Supplemantry_table_1.xlsx). As observed, without prior genotype information, our pipeline has identified 33% of SNPs and 60% of genes also detected with prior genotype information. While our pipeline identified 30% of SNPS and 77% of genes also detected with prior genotype. The remaining SNPs and genes not being detected could be attributed to data limitation, as DNAseq has more accurate information on variants, compared to RNAseq. This is because the distribution of RNAseq to reads is often non-uniform and ambiguous reads corresponding to alternatively spliced variants are often mismapped and ultimately

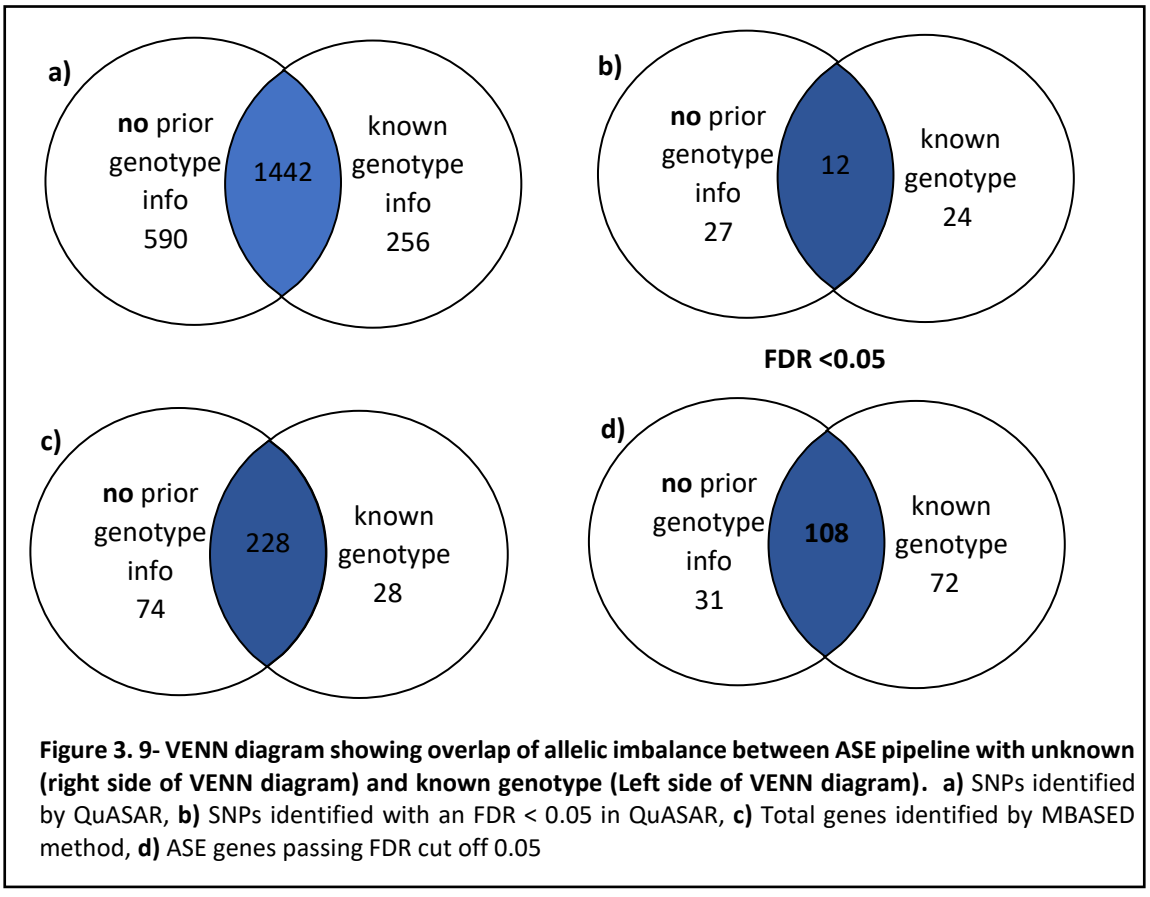


Figure 3. 9- VENN diagram showing overlap of allelic imbalance between ASE pipeline with unknown (right side of VENN diagram) and known genotype (Left side of VENN diagram). a) SNPs identified by QuASAR, **b)** SNPs identified with an FDR < 0.05 in QuASAR, **c)** Total genes identified by MBASED method, **d)** ASE genes passing FDR cut off 0.05

lost. Thus, other tools like Alleleseq identify variants from DNA-seq and phase them using Trio sequencing, while our ASE pipeline identifies variants from RNA-seq and phases them without prior haplotype information.

3.3.2.2 ASE pipeline helps to identify immunity-related genes in macrophages

By applying our ASE pipeline at gene-level mode (MBASED) to the previously described macrophage RNA-seq dataset from 169 healthy individuals and 96 *Salmonella typhimurium*-infected patients without any prior haplotype information, allelic-specific genes were detected separately for unstimulated and *Salmonella typhimurium*-infected samples. In total, 12640 genes were analysed and found to have > 1 heterozygous position in unstimulated macrophages, and 10058 in *Salmonella typhimurium*-infected samples, with 9305 being common between the two conditions. Out of these, 3150 genes had an FDR < 0.05 and major haplotype fraction > 0.6 in unstimulated macrophages, and 2828 in *Salmonella typhimurium*-infected samples. Overall, 1519 were common between the two conditions, while 1631 were present only in unstimulated samples and 1309 only in *Salmonella typhimurium*-infected samples (**Figure 3.10**).

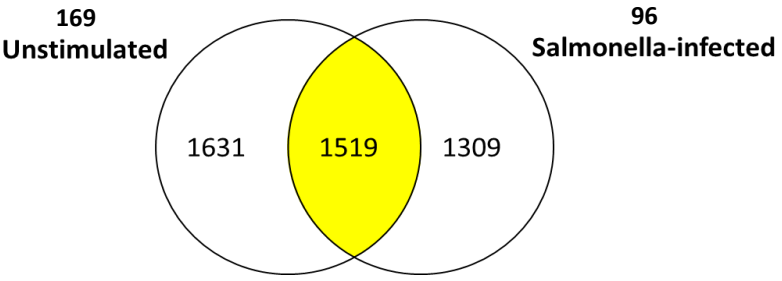


Figure 3. 10- Significant ASE genes with FDR < 0.05 in unstimulated and Salmonella-infected samples

Next, we compared our ASE genes list to 2586 eGenes (eGenes are genes whose expression levels are associated with genetic variants) identified by by S. Hellmut and colleagues through an eQTL study performed on monocyte-derived macrophages and LPS muramyl-dipeptide (MDP) treated macrophages (M1) (Kim-Hellmuth *et al.* 2017). Thus, we found 417 genes out of the 3150 ASE genes observed in unstimulated macrophages, and 409 out of 2828 seen in *Salmonella typhimurium*-infected samples. In total, there were 1519 common ASE genes between unstimulated and *S.typhimurium*-infected samples, with 255 of these being recorded as eGenes.

In order to test the statistical significance of this overlap, Fisher's exact test was applied, and the odds-ratio was calculated between the overall significant and not-significant allelic-expression genes overlapped with eGenes. The results, also shown in **Table 3.7**, recorded a p-value of 0.045 for

unstimulated macrophages, with an 0.89 odd ratio, while *Salmonella typhimurium*-infected samples had a p-value of 0.39 and an odds ratio of 0.95. The p-value for the overlapping ASE genes between unstimulated and *S.typhimurium*-infected was 0.298, with an odds ratio of 0.92. These results show there is no significant overlap between unstimulated macrophages' ASE genes with eQTLs in this set of study.

Table 3 7- Input for Fischer exact test. between Unstimulated macrophages significant FDR < 0.05 and not significant FDR < 0.05 **a)** unstimlated macrophages **b)** salmonella-infected samples **c)** common ASE genes between unstimulated and salmonella infected samples

a)

Unstimulated	Significant ASE	Non-significant ASE	Fischer exact test
eGenes	417	1394	
Not-eGenes	2733	8096	
			0.045

b)

Salmonella typhimurium	Significant ASE	Non-significant ASE	Fischer exact test
eGenes	410	1098	
Not-eGenes	2418	6132	
			0.39

c)

Common	Significant ASE	Non-significant ASE	Fischer exact test
eGenes	225	1238	
Not-eGenes	1294	6548	0.21

In order to understand whether there is a pattern of biological relevance among the genes detected, a random selection was performed for the ASE genes overlapping with eGenes. One of the ASE genes with an FDR of 0.004 is CASP8 and FADD-like apoptosis regulator (CFLAR), which is known to have a significant role in inflammatory response and fibrosis (Xiaohong *et al.* 2019); FCN1 (FDR 0), which is a member of the complement system that plays an important role in the innate immunity system, as overexpression of FCN1 in peripheral blood mononuclear cells was found to lead to Takayasu arthritis (Okuzaki *et al.* 2017); APOE (FDR 0), which codes for protein apolipoprotein E and is associated with cardiovascular diseases and Alzheimer's disease (Rowczenio *et al.* 2011); and the NLRP2 gene (FDR 0.00004), part of the nucleotide-binding family and known to significantly impact the activation of immune responses during infection and tissue injury (Komada and Muruve 2019). Genes belonging to this family play an important role in regulating the IL-1β, NF-κB production for positive regulation of inflammatory responses (Rossi *et al.* 2019). Overall, our analysis's this selection of filtered genes suggests that most have important immunity-related functions. A full list of genes is given in supplementary table 2 (https://github.com/srmeetd/ASE_supplemantry_DATA/blob/main/Supplemantry_table_2.xlsx).

3.3.2.3 miRNA binding sites are altered in ASE genes

In addition to the above analysis, we also investigated SNPs altering miRNA binding sites in ASE genes in *Salmonella typhimurium*-infected samples, which could be responsible leading to allelic imbalance. We detected 2003 SNPs across 3'UTRs of 773 ASE genes that created 1403 novel miRNA binding sites, and 2106 SNPs across 3'UTR of 805 ASE genes destroying 1355 binding sites in MDMs. Also, in *Salmonella typhimurium*-infected samples, we observed that 2342 SNPs present in 3'UTRs of 901 ASE genes created 1511 novel miRNA binding sites, while 2428 SNPs across 917 ASE genes were responsible for degrading 1416 existing miRNA binding sites.

As we can observe, there was no difference between the number of SNPs present (~2000), the number of miRNA binding sites either creater or destroyed (~1400), or the number of ASE genes with modified 3'UTRs (~700-900) in one condition over the other. This suggests that there are number of SNPs altering miRNA binding sites in ASE genes, but we cannot conclude that higher or lower number of altered miRNA binding sites leads to more or less ASE, respectively. More than that, we do not know if these miRNAs would actually have an impact on their targets just by having altered binding sites, as we do not know whether these miRNAs are in fact expressed in the two conditions.

In this sense, we next grouped all miRNAs by their expression in both M^{UN} and $M^{LPS+INF\gamma}$ (detailed explained in chapter 4) macrophages, and we divided these into three quartiles: low, medium and upper quartile. The latter set contained highly expressed miRNAs in both M^{UN} and

$M^{LPS+INF\gamma}$, which were further investigated if their binding sites in 3'UTRs of ASE genes were altered by SNPs. Overall, we found that 183 genes with only destroyed binding sites for 158 highly expressed miRNAs caused by 247 SNPs, while 218 targets had only novel binding sites created for 162 miRNAs, caused by 275 SNPs. On the other hand, 204 ASE genes were seen to have both degraded and created binding sites for highly expressed miRNAs on the same 3'UTR, with 364 SNPs destroying 193 of these binding sites and 366 SNPs creating 161 novel ones in the same 3'UTR, across the 204 targets. (**Table 3.8**). This result suggests that although genes shows Allelic imbalance, but approximately only 10% of ASE genes shows alteration of miRNA binding sites due to the variants in there 3'UTR.

Table 3 8 - Number of genes in which SNPs a) created and degraded miRNA binding sites of highly expressed miRNAs in the same 3'UTR, or b) either created or degraded miRNA binding sites.

a)	Created miRNA BS	Destroyed miRNA BS
miRNAs	193	161
SNPs	366	364
ASE genes	204	
b)	Created miRNA BS	Destroyed miRNA BS
miRNAs	162	158
SNPs	275	247
ASE	218	183

3.3.2.4 SNPs altering miRNA binding sites in 3'UTRs lead to ASE of target genes

The analysis described above was performed by selecting ASE genes (thus at gene level), and then selecting SNPs in these genes to investigate. In contrast, the following investigation was performed to identify SNPs linked to allelic imbalance (hence at SNP level), and to assess how many of these are located in and alter miRNA binding sites in 3'UTRs. This analysis was done only on RNAseq from the 169 MDMs.

All SNPs present in all genes were first filtered through the QuASAR tool integrated in our ASE pipeline, which returned a total of 81,913 SNPs across 20,136 genes, out of which 10,466 SNPs were present in 3'UTRs of 4,549 genes and were responsible for the creation of 2,120 novel miRNA binding sites. Similarly, 10,960 SNPs across 3'UTRs of 4,627 genes destroyed 2,100 miRNA binding sites. Setting a cutoff of FDR < 0.05, we found 644 SNPs present in 3'UTRs of 261 genes that showed allelic imbalance (linked to ASE), out of which 81 SNPs created 130 novel miRNA binding sites in 80 genes.

Similarly, 155 SNPs in 3'UTRs of 128 genes destroyed 220 miRNA binding sites, taking into account the direction of the beta score (**Box 3.1**).

Box 3.1: Beta Score

The beta score shows the strength and direction of ASE, with scores > 0 showing bias towards the reference allele, and vice versa for < 0. Creating a new miRNA binding site would mean reduced expression of the mutant variant and thus, its associated beta score would be negative. Otherwise, if the existing miRNA binding is degraded, the alt allele is considered protected, and the beta score would be positive. What is more, a beta score of 0 indicates there is no change in the SNP, and it can be considered a false-positive result. Therefore, we rejected any SNPs with a positive beta score that created new miRNA binding sites and any SNPs that degraded existing binding sites and had a negative beta score. This method offered high confidence in detecting SNPs altering miRNA binding sites and being responsible for the allelic imbalance.

Furthermore, upon investigating SNPs linked to ASE (with an FDR < 0.05) that were present in and altered miRNA binding sites in 3'UTRs of previously identified ASE genes in MDMs, we found 644 SNPs linked to allelic imbalance from QuASAR , out of which 83 SNPs were detected in 3'UTRs of 43 ASE targets and disrupted 122 existing miRNA binding sites, while 39 SNPs created 61 novel binding sites in 25 ASE genes. However, none of these SNPs destroyed or created binding sites for miRNAs highly expressed in M^{UN} samples, but the expression of miRNAs whose binding sites were altered in ASE genes was discussed in the above section 1.4.1

In addition, we also looked at the highest number of SNPs linked to ASE that were present on single ASE genes, irrespective of their alteration of miRNA binding sites. Thus, we observed that the highest number of SNPs linked to ASE by QuASAR, and placed on ASE targets identified by MBASED method were 61 in the 3'UTR of HLA-DQB1 ASE gene, followed by 44 SNPs in the 3'UTR of HLA-DQA1 and 18 on HLA-DRB1's 3'UTR (**Table 3.9**). However, HLA genes have been previously found to have high genomic variations (T Shiina *et al.* 2009), including in their 3'UTRs (A Sabbagh 2014), hence these were eliminated from our further analyses, in order to avoid biases. In addition to that, we also disregarded ASE genes with 0 SNPs in their 3'UTRs, as we were only interested to investigate the potential correlation between SNP-derived variations and allelic imbalance.

Thus, upon filtering out HLA genes, we found 17 SNPs in the 3'UTR of XIAP gene, followed by 11 SNPs in the 3'UTR of GREM1 and MDM2, and 7 SNPs in APOL1 and SERPINB9 genes. All these genes have been previously shown to have variations leading to allelic imbalance (M M Gerber *et al.* 2012; J

Heighway *et al.* 1994), thus increasing the confidence in our detection of SNPs and genes linked to ASE. We observed that a minimum of 1 SNP was identified in 85 ASE genes, 2 SNPs were present in 27 ASE genes, and a maximum of 17 SNPs were present in just one ASE target, the XIAP gene (**Table 3.9**). Therefore, we can observe that the number of ASE genes with increasing number of SNPs is drastically reduced even starting from just 2 SNPs present. This result suggests that as small as 1 variant could be responsible in allelic specific expression of genes.

Table 3 9- Number of SNPs linked to ASE detected in different number of ASE genes

No. of SNPs	Number of ASE genes
1	85
2	27
3	13
4	10
5	5
6	4
7	2
11	2
17	1
18	*1*
44	*1*
61	*1*

3.3.2.5 ASE genes in M1 phenotype are enriched in inflammatory pathways

In order to observe whether ASE genes are involved in macrophage polarisation, we further overlapped the 3,150 ASE genes identified in 169 MDMs (controls), and the 2,175 ASE genes found in 96 *Salmonella typhimurium*-infected samples (**Figure 3.10**) with 5,681 differentially expressed genes between 8 M^{UN} and 8 $M^{LPS+INF\gamma}$ from RNAseq data generated by our lab. We found that out of 3,150 significant ASE genes in MDMs, 833 were differentially expressed between M^{UN} and $M^{LPS+INF\gamma}$ (FDR < 0.05), with 473 being downregulated and 360 being upregulated in $M^{LPS+INF\gamma}$. Similarly, out of 2175 ASE genes detected in *Salmonella typhimurium*-infected samples, 601 were differentially expressed between M^{UN} and $M^{LPS+INF\gamma}$, 272 being downregulated and 329 upregulated in $M^{LPS+INF\gamma}$.

Furthermore, gene enrichment analysis showed that ASE genes obtained from MBASED from *Salmonella typhimurium*-infected samples that were also upregulated in $M^{LPS+INF\gamma}$ were mainly involved in inflammatory-related biological function (**Figure 3.11**). This result indicates that genes playing

important roles in maintaining and/or regulating immune responses within cells may have allele-specific expression. In contrast, the ASE genes in MDMs that were upregulated in M^{UN} compared $M^{LPS+INF\gamma}$ treatment to did not seem to be enriched in any particular biological processes. We observed the same for downregulated ASE genes, which also did not return any particular GO terms related to any specific functions.

Moreover, to understand the function of the genes for which SNPs in their 3' UTR were linked to allelic imbalance and were predicted to alter miRNA binding sites, we performed Gene Ontology (GO) enrichment analysis (**Figure 3.12**). Thus, we observed that these genes are mainly involved in signalling, inflammatory-related and apoptosis-related pathways. Similarly, we also investigated the relationship between SNPs showing ASE and altering miRNA binding sites with coronary heart disease (As discussed in introduction SNPs in TRIB1 have been identified to bed associated with coronary heart disease) identified by GWAS analysis. However, we found that none of the SNPs were directly linked

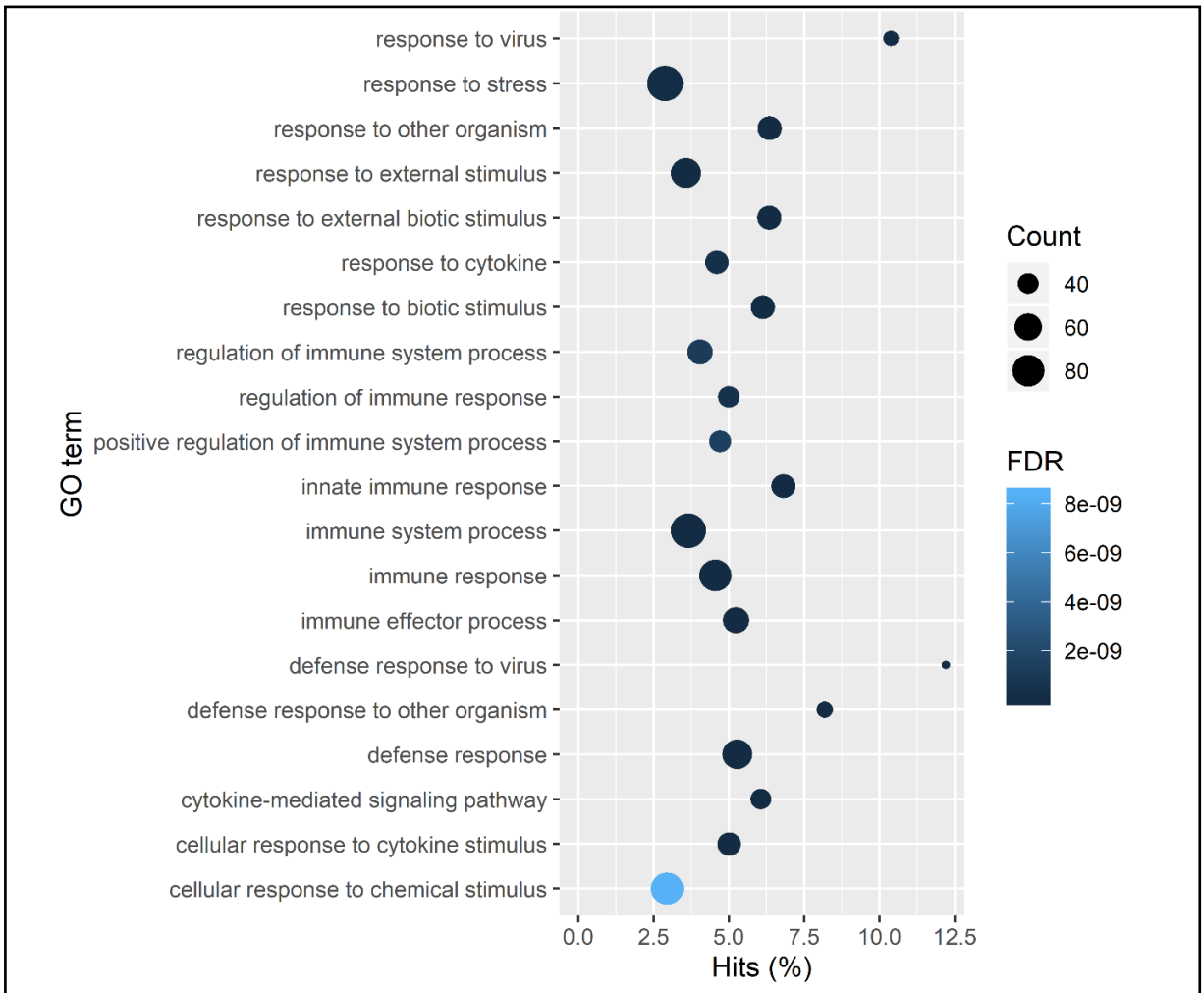


Figure 3. 11- Top 20 GO terms observed for ASE genes in Salmonella typhimurium-infected samples and upregulated in MLPS+INFγ (M1 phenotype). Dot plot represents top 10 gene ontology annotations, x-axis shows the percentage of differentially expressed genes present in particular ontology function, y-axis represents gene ontology annotations, dot size represents the number of differentially expressed present in corresponding gene onotology function and color represents the FDR score.

with coronary heart diseases, but some SNPs were in linkage disequilibrium with the SNPs associated with several other GWAS diseases/trait (**Table 3.10**).

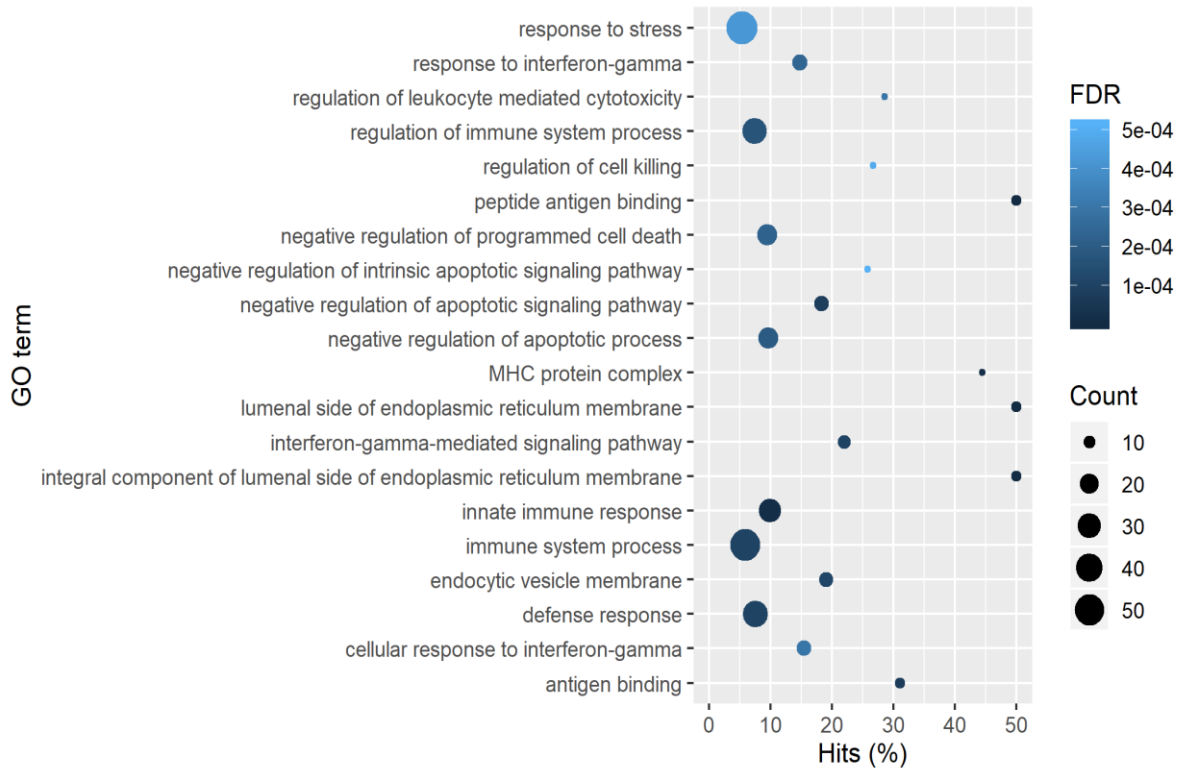


Figure 3. 12- Gene Ontology for genes with SNPs linked to ASE. Dot plot represents top 10 gene ontology annotations, x-axis shows the percentage of differentially expressed genes present in particular ontology function, y-axis represents gene ontology annotations, dot size represents the number of differentially expressed present in corresponding gene onotology function and color represents the FDR score.

Table 3 10- List of SNPs in linkage disequilibrium with SNPs associated with GWA study; Query – ASE SNPs altering miRNA binding sites, GWAS Trait – Disease name, RS number – linkage disequlibrium SNPs associated with disease trait, R^2 – R square a measure of correlation of alleles of query, RS – number where 0 means alleles are independent and 1 means allele in one variant predict the allele in another variant , D' – an allelic segration between two variants (Query and Rs-number), values ranging 0-1, higher value indicates tight linkage of alleles, Risk allele – Minor allele frequency, Effect Size – Effect size 95% confidence interval, P-value – P-value of significance obtained from GWAS

Query	GWAS Trait	RS Number	R2	D'	Risk Allele	Effect Size (95% CI)	P-value
rs1513890	Coronary artery calcified atherosclerotic plaque in type 2 diabetes	rs6829172	1	1	NR	0.347	8.00E-06
rs17162846	Systolic blood pressure	rs1014988	0.6158526	0.994812	0.4781	0.2039	2.00E-09
rs17162846	Electrocardiogram morphology	rs17162854	0.608614973	0.969478	0.500138	0.0304694	4.00E-08
rs17162846	Diastolic blood pressure	rs7546498	0.614363109	0.989666	NA	0.147	5.00E-10
rs17162846	Systolic blood pressure	rs7546498	0.614363109	0.989666	NA	0.257	3.00E-11
rs13427770	Colorectal cancer	rs11903757	0.688274016	0.992438	0.16	1.16	4.00E-08
rs1050450	Coronary artery disease	rs11718165	0.921396736	0.986141	NR	0.0327978	6.00E-08
rs1050450	Chronic inflammatory diseases	rs3197999	0.875274178	0.985657	NR	NA	7.00E-55
rs8005	Triglycerides	rs709822	0.943841228	0.99026	0.7117	0.0212	5.00E-13
rs8005	Triglyceride levels	rs3947	0.816313807	1	0.25	0.022	3.00E-09
rs2287367	Prostate cancer	rs12634	0.946956188	1	NR	1.058	4.00E-14
rs2287367	Prostate cancer (early onset)	rs1182	0.881708292	1	NR	1.13	1.00E-07
rs2287367	Prostate cancer (advanced)	rs1182	0.881708292	1	NR	1.09	1.00E-07
rs582452	HDL cholesterol	rs56959712	0.792315372	0.979026	0.2159	NA	7.00E-13
rs582452	HDL cholesterol levels	rs113740515	0.78410937	0.965175	0.79055	0.0377999	1.00E-58
rs582452	Triglyceride levels	rs580063	0.781039957	0.972035	NR	0.0233	4.00E-13
rs582856	HDL cholesterol	rs56959712	0.798504657	0.992891	0.2159	NA	7.00E-13
rs61955033	HDL cholesterol levels	rs2454703	0.979816277	0.991867	NR	0.005388	4.00E-13
rs1060314	Systemic mastocytosis	rs9937881	0.880489696	1	0.2089	2.286	3.00E-08
rs3211567	White coat effect	rs2292954	0.888128296	0.970139	0.11	6.2	2.00E-06
rs3211567	Type 2 diabetes	rs12932337	0.970640273	0.985211	0.8301	0.034	8.00E-11

3.4 Discussion

In this chapter, we investigated the post-transcriptional regulation of TRIB1, especially SNPs in its 3'UTR responsible for altering miRNA binding sites, and whether or not these SNPs are linked to allelic imbalance. We found that sequence variants existed in TRIB1 3'UTR and that several of these alter miRNA binding sites; for example, rs622521034 SNP created a novel binding site for miR-29-3p. Furthermore, SNPs in TRIB1 3'UTR were associated with traits by GWAS. Our ASE pipeline was successfully validated and we observed that miRNA binding sites are altered in ASE genes, that SNPs linked to allelic imbalance alters miRNA binding sites in 3'UTRs, and that ASE genes in M1 phenotype of macrophages are enriched in inflammatory-related pathways.

Overall, we found 60 variants in 3'UTR of TRIB1, out of which at least one was detected in 156 MDM control samples. The maximum of variants was identified in one patient TRIB1 was 13, and the minimum was one. More than that, we observed that the number of TRIB1 3'UTR variants increased with an increase in their expression level (**Figure 3.6**), which could be due to SNPs destroying miRNA binding sites that would downregulate its expression. Upon investigating the alteration of such miRNA binding sites by SNPs in 3'UTR of TRIB1, we found that 11 variants degraded 20 binding sites across 24 samples, while 18 SNPs were responsible for creating novel miRNA binding sites for 39 samples (**Figures 3.5** and **3.7**).

In order to validate our computational analyses, we chose one SNP, rs62521034, that created a novel miRNA binding site for miR-29a-3p. The variants containing this SNP rs62521034 led to a lower expression of TRIB1, as measured by luciferase reporter, when miR-29a-3p was overexpressed, confirming that the creation of a binding site for this miRNA enabled its targeting of TRIB1 (**Figure 3.7**). These results suggest that a variant in TRIB1 created a new miRNA binding, which reduces the expression of TRIB1. Moreover, it confirmed that our pipeline successfully identifies putative miRNA binding site changes. However, considering bioinformatics analyses detect a large number of candidate SNPs, more examples would need to be experimentally validated in order to test and prove the accuracy of our strategy.

Out of all the miRNAs whose binding sites were altered by SNPs in 3'UTR of TRIB1, only nine were expressed in both unstimulated and M1-like small-RNA-seq datasets. Six of these miRNAs had novel binding sites created in the 3'UTR of TRIB1, while three had existing sites degraded. However, among the nine miRNAs, only one miR-29a-3p was differentially expressed between the two conditions, M^{un} and $M^{LPS+IFN}$, and upregulated in $M^{LPS+IFN}$, which could be one of the reasons for the low expression of TRIB1 in M1-like macrophages; moreover, it was also a reason for validating this miRNA in HEK293T cell line by creating a mutant in 3' UTR of TRIB1. Overall, we observed no difference

between the number of miRNA binding sites created and destroyed by SNPs in 3'UTR of TRIB1. However, the SNPs detected may regulate the expression of TRIB1 by altering the binding of other MRE factors, such RNA-Binding proteins (RBPs), translational repressors, splicing factors, and riboswitches, or in other cell types.

Our analyses described above showed that SNPs in 3'UTRs are responsible for altering miRNA binding, which we then investigated further whether these may also lead to allele-specific expression (ASE). In this sense, we developed our own cgat-based ASE pipeline, which was designed to accept inputs of RNA-seq, DNA-seq and Chip-seq data sets as both unaligned and aligned files. The advantage of our ASE pipeline over other available tools/frameworks is that it does not rely on prior known phasing or trio sequencing for detecting ASE events, being capable of identifying variants and then phasing them on its own. More than that, it also deals with sequencing errors by generating parental genomes from the phased variants as it helps to differentiate between real varaints and sequencing error. Despite being able to detect variants and ASE events even without the need of any prior haplotype information, our pipeline has been tailored to also accept and use this, when available. Other important features of our pipeline are listed below:

1. RNA editing events can be filtered when RNAseq files are provided as input
2. It calculates Minor allelic frequency for each sample set
3. It offers a threshold option for Hardy-Weinberg's equation (which considers that "genotype frequencies in a population remain constant between generations in the absence of disturbance by outside factors") (Ryckman and Williams 2008).
4. Identifies ASE on both SNP and gene levels
5. It detects whether ASE SNPs are responsible for altering miRNA binding sites

For validating our pipeline, we applied it to the NA12878 RNAseq data, with and without known genotype and haplotype information (derived trio DNA seq of both the sample, and its parents), and we recorded a 1.98% increase in mapping percentages when no prior haplotype information was used (**Table 3.6**). This increase may be due to reads being mismapped to SNPs location identified by GATK, as the pipeline relies on predictions when not using prior genotype information. All in all, these confirm the success of our pipeline in generating results without the need of prior genotype information, comparable to those obtained with known genotype information, as approximately 60% of ASE genes detected with known genotype information were also accurately identified with no prior information and 77% of ASE genes detected with no prior information (**Figure 3.9**). However, MBASED was used in both cases, and these values are to be considered valid only to the extent of how accurate MBASED is.

Out of a total of 3150 significant ASE genes (FDR < 0.05) detected in MDM control samples and 2828 in *Salmonella typhimurium*-infected samples (**Figure 3.10**), only 417 and 225 ASE genes, respectively, were found to be eGenes (**Table 3.7**). In contrast, there was a higher overlap between non-significant ASE genes and eGenes set observed, which could be because the ASE analysis considers genetics driven changes in expression levels of alleles within individuals, while eQTL mapping considers these changes between individuals. More than that, eQTL analysis counts all RNA reads from genes, whereas the ASE analysis considers only reads mapping to SNP regions. Another reason could also be that one allele may be higher expressed than the other, which leads to a feedback mechanism limiting the expression of the highly expressed one. Thus, total gene expression is then recorded as similar, irrespective of SNP genotype. Such genes are not linked to allelic imbalance in ASE analysis, but they may be identified as eGenes *via* eQTL mapping.

Furthermore, we identified a large number of SNPs altering miRNA binding sites in 3'UTRs of ASE genes in both control MDMs and *Salmonella typhimurium*-infected samples, but no difference in either the number of SNPs, miRNA binding sites nor ASE genes between the two conditions. While our observations do suggest that altered miRNA binding sites can be linked to allelic imbalance, we could not conclude whether a higher number of these lead to an increase in ASE genes. Moreover, the degradation or creation of a miRNA binding site is not enough to suggest that those particular miRNAs would indeed bind and affect their targets, as these miRNAs may not even be expressed in certain conditions. In this sense, we looked at the expression of miRNAs in the two respective macrophage polarisation states (M0 and M1) and we observed that the number of ASE genes with altered binding sites for highly expressed miRNAs was lower than those reported above, without taking expression into account. However, we detected both ASE genes with 3'UTR variants with either degraded or created binding sites, as well as ASE genes with 3'UTRs containing both degraded and created miRNA binding sites (**Table 3.8**).

Next, we observed that 85 ASE genes had only 1 significant SNP identified by QuASAR in their 3'UTR, and only 1 gene had as many as 61 significant SNPs (**Table 3.9**). These results suggest that a minimum of 1 SNP could be responsible for the allelic imbalance of gene. Moreover, the number of SNPs, altered miRNA binding sites and ASE genes was further reduced when looking only at individually significant SNPs and significant ASE genes (FDR < 0.05), as identified by QuASAR. Thus, we found that out of 3150 significant ASE genes, only 43 had one of 83 significant SNPs that created 61 novel miRNA binding sites, and only 25 ASE genes presented 39 significant SNPs that destroyed 122 miRNA binding sites. Moreover, this filtering step improved our confidence in predicting SNPs and genes linked to allelic imbalance, especially through the impact on miRNA binding sites, as genes with

the maximum number of significant SNPs identified have already been found to show allele-specific expression.

Despite the multiple genes and SNPs we identified and predicted to be linked to ASE, there were no significant SNPs in the 3'UTR of TRIB1 that could be linked to ASE, not even SNP rs62521034, whose reads mapped approximately equal in both reference and alternate alleles (ref allele – 26 reads and alternate allele – 27 reads mapped). The effect of this SNP was also tested *in vitro* in HEK293T cells, but it did not lead to any significant difference in TRIB1's expression in MDMs. This could be because miR-29a-3p, whose binding site was altered by this SNP, was too lowly expressed in M^{un} . More than that, the expression of TRIB1 itself was found to be low in both MDM controls and *Salmonella typhimurium*-infected samples, which are similar conditions to M^{UN} and $M^{LPS+INF\gamma}$. This is not the case in the test in HEK293T cells, where both target and miRNA are over expressed.

In conclusion, we have developed a successful pipeline for detecting thousands of SNPs from RNAseq data sets and for identifying allelic-specific events at both gene and SNP levels. More than that, the detection of SNPs without prior genotype information was compared to those found using known genotype information from DNA-seq, and we identified a large number of SNPs from our RNAseq datasets, previously reported by DNA sequencing, as well (e.g. for TRIB1 we found 90% of SNPs previously annotated). However, one limitation of our pipeline is that it does not take into consideration multi-allelic sites and indels, and currently, only the miRanda algorithm is implemented for the detection of miRNAs targeting the binding sites altered by the SNP identified. Incorporating other tools or scanning more databases, such as TargetScan, could improve the confidence and accuracy of our ASE pipeline. Despite this, we consider our pipeline to be the first complete framework to detect allelic imbalance at both gene and SNP levels using solely RNAseq data, as well as identify altered miRNA binding sites. Moreover, our pipeline has improved power of interpretation to detect ASE genes and SNPs as compared to traditional ASE method.

Chapter 4

Role of miRNAs in macrophage polarisation

4.1 Introduction

Inflammation is an immunity response triggered by the body in fighting pathogen infections and healing physical injuries (Mantovani *et al.* 2013), and it involves the recruitment of multiple components, such as macrophages (described in detail in Introductory Chapter), which play an important role in host defense mechanism and tissue repair (Curtale, Rubino, and Locati 2019). The inflammatory response mediated by macrophages is initiated by the recognition of microbial structures known as pathogen-associated molecular patterns (PAMPs), or various endogenous signals from damaged tissues or cells, known as a damaged-associated molecular pattern (DAMPS). This is enabled through the activation of germline-encoded pattern-recognition receptors (PRRs), such as Toll-like receptors (TLRs) (Chen *et al.* 2018), leading to the production of chemokines, cytokines, reactive oxygen and nitrogen species, which help trigger pro- or anti-inflammatory responses (Turner *et al.* 2014).

The efficiency of inflammatory responses mediated by macrophages depends on the coordination of proteins involved in macrophage polarization, whose expression is regulated at the transcriptional and post-transcriptional level (Chen *et al.* 2018). miRNAs are a class of such mRNA regulatory elements (MREs) involved in gene regulation, which have been extensively researched in order to understand their importance and role in immune and inflammatory responses (Liu and Abraham 2013). Most of these studies have been performed in the context of cancer and metabolic diseases (Essandoh *et al.* 2016), but none of these has considered the influence of alternative polyadenylation (APA) on miRNA binding sites.

In eukaryotes, many genes often have more than one polyadenylation sites, which leads to multiple isoforms being generated by cleavage at different location, a process known as alternative polyadenylation (APA). It frequently occurs in 3'UTRs of mRNA transcripts as a final step in mRNAs maturation, through the cleavage of 3' UTR at polyadenylation sites of pre-miRNAs and the addition of a poly-A tail. This is important for the stability, localization, and translation efficiency of mature miRNAs, and I can also lead to further changes in mRNAs half-life or translation, by creating or destroying miRNA binding sites. (Tian and Manley 2016).

In this current project, we present a first-of-its-kind project to investigate the link between miRNAs and macrophage polarization, through transcriptomic and miRNA studies. This study will shed

light on miRNAs and their target genes involved in macrophage polarization and the effect of the APA process in regulating these genes. Furthermore, our study aims to identify differentially expressed (DE) hub miRNAs targeting more than the average number of differentially expressed genes, between unstimulated (M^{UN}) and LPS+INFγ ($M^{LPS+INFγ}$) challenged macrophages.

4.2 Materials and Methodology

Declaration

Dr. Chiara Niespolo performed all *in vitro* work under the supervision of Prof. Endre Kiss-Toth (IICD, University of Sheffield, Sheffield, UK) and I performed all the computational analysis and interpretation of results.

4.2.1 Isolation and culture of human monocyte-derived macrophages (MDMs)

Peripheral blood mononuclear cells (PBMCs) were isolated from venous blood from healthy donors by Ficoll–Paque Plus (GE Healthcare) density centrifugation; CD14+ monocytes were selected by positive magnetic separation using CD14 human microbeads (Miltenyi Biotec). Monocytes were then cultured in complete media for 7 days: RPMI-1640 (Gibco), 10% (v/v) low–endotoxin heat-inactivated FBS (PanBiotech), 1% (v/v) L–glutamine (Gibco) and 1% (v/v) penicillin/streptomycin (Gibco). 100 ng/mL of human recombinant M–CSF (Peprotech) was added to the media to differentiate monocytes into macrophages (monocyte-derived macrophages, MDMs).

4.2.2 Macrophage polarisation

After 7 days of isolation and differentiation, MDMs were treated with 100ng/mL LPS (Serotype R515 TLR grade TM, Enzo Life Sciences) and 20ng/mL of human recombinant INF-γ (Prepotech) to generate M1-like phenotype, and 20ng/mL of human recombinant IL-4 to generate M2-like macrophages. Cells were collected after 24 hours for RNA extraction.

i) Transient transfection

Transient transfections were carried out on MDMs on day 7 post isolation and differentiation. A total of 50nM of miR-155-5p, miR-125a-3p, miR-186-5p previously synthesised mimics and non-targeting control (Horizon Discovery) were transiently transfected using Viromer Green (Lipocalyx, Cambridge Bioscience), according to manufacturer's instructions. Cells were collected 24 hours post transfection for RNA extraction.

ii) Total RNA extraction

Total RNA was isolated using the miRNeasy Mini Kit (QIAGEN). RNA concentration and purity were assessed by NanoDropTM Spectrophotometer (A260/A280 ratio= 1.8-2.0) (ThermoFisher Scientific). Before the small-RNA sequencing and the mRNA sequencing, RNA integrity and concentration were also determined using Agarose Gel Electrophoresis (Gel Con: 1%; voltage: 180v; Run Time: 16min) and Bioanalyzer Agilent 2100 (RNA Integrity Number ≥ 7.5).

iii) Real-time quantitative PCR (RT-qPCR)

cDNA synthesis and RT-qPCR were carried out using the miRCURY LNA RT Kit and miRCURY LNA miRNA PCR Assay Kit, based on SYBR Green chemistry (QIAGEN) manufacturer's instructions. Results were analyzed upon a CFX384 C1000 Touch Thermal Cycler (Biorad), using the 2-ΔCt method. PCR primers for miR-155-5p, miR-125a-3p, miR-186-5p, miR-149-5p, miR-1343-4p, miR-766-3p were purchased from QIAGEN and diluted according to manufacturer's instructions.

iv) Viability assay

Viability assay was carried out on supernatants collected from MDMs, 24 hours post-transfection with miRNA mimics/control, using the kit RealTime-Glo™ MT Cell Viability Assay (Promega) manufacturer's instructions.

4.2.3 Small-RNA and mRNA sequencing

The genome sequencing company Novogene Co. Ltd (https://en.novogene.com) performed small RNA for the 8 unstimulated (M^{UN}) and 8 M1-like macrophages ($M^{LPS+INF\gamma}$), using the Illumina platform with single-end 50 bp and RNA-seq for 6 for hubm-RNA (miR-125a, miR-155,miR-186 and control) using the illumunia platform with $20M$ paired-end 150 bp.

4.2.4 Macrophage polarized small-RNA-seq bioinformatics analysis

i) Quality check and mapping

The quality of reads was checked by FastQC implemented in cgat-pipeline readqc (https://github.com/cgat-developers/cgat-flow/blob/master/cgatpipelines/tools/pipeline_readqc.py). Reads with poor quality (reads with quality score <30) were removed and adapter sequences were trimmed using cutadapt with default parameters, implemented in cgat-pipeline readqc (https://github.com/cgat-developers/cgat-flow/blob/master/cgatpipelines/tools/pipeline_readqc.py) (Supplementary Figure 2). Clean reads were mapped against human small-RNA sequences downloaded from the RNAcentral database (Sweeney *et al.* 2019), using bowtie2 (Langmead *et al.* 2009) local alignment option with default parameters. RNAcentral was used instead of the whole human genome sequence (hg38), because it

contains only small-RNA sequences, while hg38 also includes information on protein coding genes. Hence, small-RNA reads of approximately 50 base pairs could map to multiple locations in hg38, including to protein coding sequences, and may produce false-positive results, while by using RNAcentral, these were only mapped to small-RNA sequences. Raw counts were then obtained from samtools (Li and Durbin 2009). The RNA-central database was also used to annotate small RNAs, and only miRNAs were considered for further analysis.

ii) Principle component analysis

To observe whether samples are clustering based on different conditions, we performed principal component analysis (PCA). To generate PCA components, we passed rlog normalized data through DESeq2 R package to procomp R package. Next, the ggplot2 R package was used to visualize components 1 and 2 of PCA.

iii) Differential expression analysis

Differential expression analysis was performed between M^{UN} and $M^{LPS+INF\gamma}$ using the DESeq2 R package. miRNAs with < 0.05 FDR and log fold change cutoff |≥1| were considered for further analysis, they were further visulzied in heatmap generated using Complexheatmap R package. Target genes of differentially expressed miRNAs were identified using the TargetScan (Lewis, Burge, and Bartel 2005) (https://github.com/srmeetd/miRNA-mRNA_integrative_analysis/blob/main/small_RNA_modified.Rmd).

4.2.5 Macrophage polarized mRNA-seq bioinformatics analysis

i) Dataset

A previously published RNA-seq dataset (Baidžajevas *et al.* 2020) from our laboratory was re-analyzed, focussing on differential gene expression between M^{UN} and $M^{LPS+INF\gamma}$ conditions. The dataset included 16 samples, including 8 M^{UN} and 8 $M^{LPS+INF\gamma}$ (resulting in an M1-like phenotype).

ii) Quality check, mapping and differential expressed analysis

Reads were checked for quality using FastQC software implemented in cgat-pipeline readqc . Low-quality reads were removed and adapters were trimmed using trimmomatic algorithms with default parameters using the readqc pipeline. Samples were further processed and quantified with Salmon (Patro *et al.* 2017), using the Ensemble 93 human reference transcriptome. Next, R package tximport (Soneson, Love, and Robinson 2015) to generate gene level expression. Differential expression of genes was conducted using DESeq2 (Love, Huber, and Anders 2014) R package, and genes with an FDR < 0.05 and logFC cutoff |>1| were selected for further analysis

((https://github.com/srmeetd/miRNA-mRNA_integrative_analysis/blob/main/small_RNA_modified.Rmd).

iii) Putative target identification of miRNAs

In order to investigate the function of differentially expressed miRNAs, we selected their corresponding target mRNAs with only 6-8 seed sequences of 74 conserved and non-conserved miRNAs, using TargetScan database. The predicted target mRNAs were further filtered and only those that were differentially expressed in the mRNA transcriptome data were retained for further analysis.

iv) miRNA-mRNA integrated analysis

Pathway and gene ontology enrichment analysis was performed using GOseq (Young *et al.* 2010) R package. First, we extracted 3'UTR sequences from hg38 GTF file, calculated their length and then removed length biased 3'UTRs. Next, names of KEGG pathways were identified using KEGGREST (Tenenbaum 2020) R package. Further, FDR was calculated on the 'over_represented_pvalue' column output obtained from GOSeq. The reason for using the 'over_represented' column is that it represents more DE genes in the present categories than expected at any given size of category and length distribution; hence, these are considered enriched DE genes in that category (Young *et al.* 2010). Pathways passing the threshold of FDR < 0.05 were then selected for further analysis.

Common target genes between up and downregulated sets of miRNAs were further identified by using the Jaccard index. Jaccard index represents the number of target genes common in pairwise miRNAs over the total number of target genes of a miRNA.

Alternative polyadenylation (APA) analysis

APA site usage was estimated by identifying changes in 3'UTR lengths of a gene between M^{UN} and $M^{LPS+INF\gamma}$ macrophages. The DaPars (Xia *et al.* 2014) method with the following parameter was used to identify the genes whose 3'UTR length changed between the two conditions.

```
Num_least_in_group1=1 \
Num_least_in_group2=1 \
Coverage_cutoff=30 \
FDR_cutoff=0.05 \
PDUI_cutoff=0.5 \
Fold_change_cutoff=0.59 \
```

4.3 Results

In order to investigate the role of small RNAs (miRNAs) in MDMs in pro-inflammatory M1 polarisation states, we conducted small RNA seq on 8 M^{UN} and 8 $M^{LPS+INF\gamma}$ samples, as previously described. The analysis results showed that more than 70% of the sequences recovered mapped to miRNAs (**Appendix table 6**), with ribosomal RNAs being the next largest single category (**Figure 4.1a**). Looking at miRNA expression levels separating treated from untreated samples on the first two principal components, (**Figure 4.1b**) we found that 102 miRNAs were least two-fold up or downregulated, with a 5% false discovery rate (**Figure 4.1c,** supplementary Table 1 https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantery_table1.xlsx). These differentially expressed miRNAs were very clearly separated between the stimulated from unstimulated macrophages (**Figure 4.1d**).

Furthermore, to study the effect of these changes in miRNAs expression on their target genes, we performed differential expression analysis on mRNA-seq data of a separate set of 8 M^{UN} and 8 $M^{LPS+INF\gamma}$ samples (Alignment: **Appendix table 7**). Similar results as above were recorded, with mRNA expression profile also being clearly separated between stimulated and unstimulated macrophages, with 5,681 genes being identified at least two-fold up or downregulated, at a 5% FDR threshold (**Figure 4.2,** supplementary Table2, https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantery_Table2.xlsx).

Thus, these results show that pro-inflammatory stimulation is accompanied by a major shift in both the long (mRNA) and short (miRNA) RNA transcriptomes of primary human macrophages.

4.3.1 Differentially expressed miRNAs preferentially target genes involved in pro-inflammatory pathways

By investigating the role of 102 differentially expressed miRNAs in regulating macrophage polarisation to and between M0 and M1 states, we identified a number of miRNAs whose target genes recorded a significant change in expression, but in a different direction of that for miRNAs, upon LPS+INFγ stimulation (e.g. downregulated targets of upregulated miRNAs and vice-versa). Thus, we found 1531 upregulated putative target genes of 44 downregulated miRNAs, and 1767 downregulated putative target genes of 26 upregulated miRNAs.

Furhermore, KEGG and Gene Ontology (GO) enrichment analysis helped us understand more about the biological function of differentially expressed miRNAs on their predicted mRNA targets. Thus, we observed that significantly upregulated targets of downregulated miRNAs were involved in

pathways responsible for immune responses and activation of inflammatory-related cascades, such as MAPK signalling, in cytokine and chemokine signalling, the JAK/STAT pathway, toll-like receptor (TLR) signaling pathway, as well as apoptosis and T-cell receptor signalling pathways. In contrast, targets of upregulated miRNAs are enriched for genes involved in the cell cycle and DNA replication (**Figure 4.3,** supplemantry table 3 https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantary_data3.xlsx).

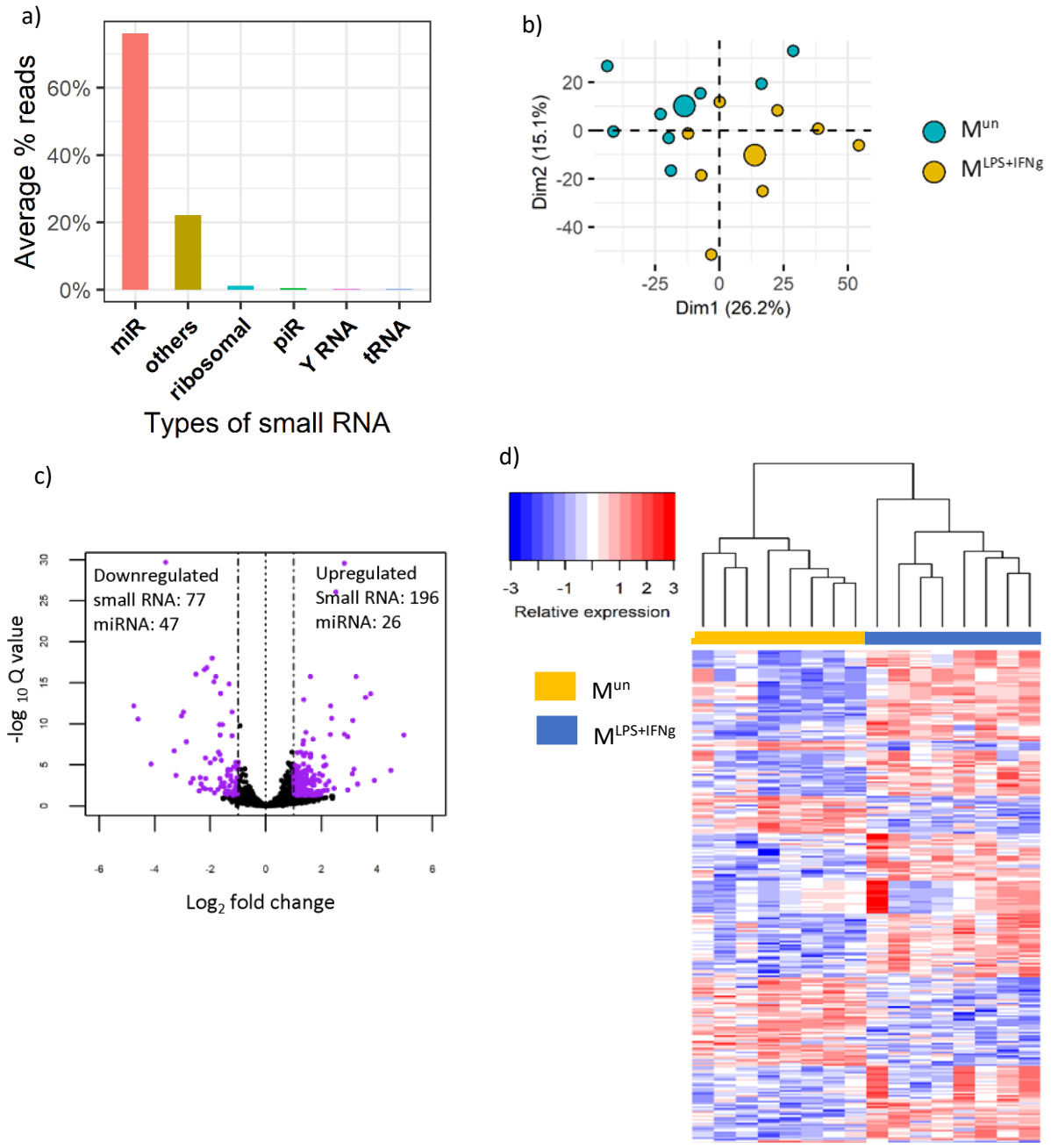


Figure 4. 1- Small non-coding RNA seq in human polarised macrophages: Mun vs MLPS+IFNγ a) Average percentage of reads mapping to different small RNAs classes, from the small non-coding RNA-se; x-axis represents different catgories of small RNAs: miR – miRNAs, ribosomal - ribosomal RNA, piR - Piwi-interacting RNA (piRNA), Y RNA - Y RNA; tRNA - transfer RNA; others - lncRNA, long non-coding RNA, small nuclear RNA, antisense small RNA; y-axis shows average percentage of small RNAs across libraries. **b)** PCA plot of small RNA seq based on log2 (CPM+1) values; dots represent different samples, large dot represents the centroid position **c)** Volcano plot showing differentially expressed RNAs (purple) (FDR<0.05, abs(log2 Fold Change>1)) **d)** Heatmap of differentially expressed miRNAs in M0 and M1 macrophages, illustrating the relative expression levels and clustering for two genotypes (MO and M1-like macrophages). Each colored cell in the heat map represents the standardized relative gene expression value for samples in each condition. The largest gene expression values are displayed in red color, intermediate values in shades of red and blue and the smallest values in darkblue.

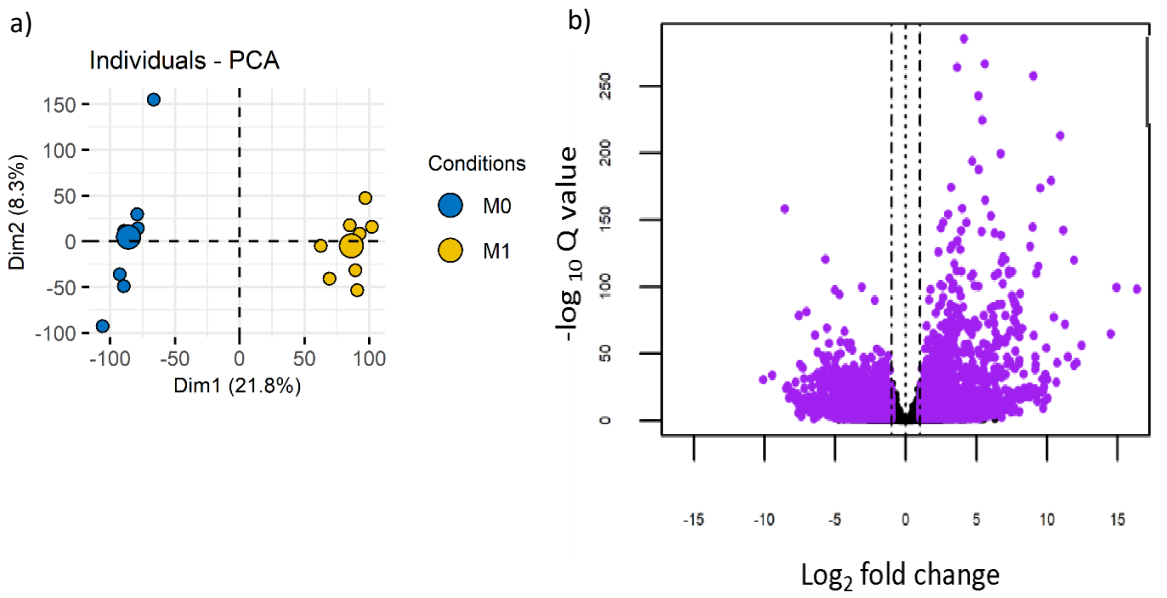


Figure 4. 2- mRNA seq in human polarised macrophages: Mun vs MLPS+IFNγ. a) PCA plot of mRNA seq based on log2 (CPM+1) values; dots represent different samples, large dot represents the centroid position . ***b)*** **Differential gene expression** with fold expression difference (x-axis) between log normalized expression values in 8 M^{un} samples and 8 $M^{LPS+IGNg}$ samples, plotted version -log_{10} FDR value (y axis). Purple dots represent the genes showing significant differential expression selected by using dual threshold of FoldChange and FDR values.

The above analysis results suggest that miRNAs may play an important role in the cell cycle progression, and may also enable the production of inflammatory cytokines and activation of immune responses in macrophages challenged by pathogen infections. Similar patterns were observed through Gene-ontology analysis (**Figure 4.4**), which showed that upregulated target genes of downregulated miRNAs showed an enrichment in regulators of cytokine/chemokine receptor activity. In contrast, the downregulated targets of upregulated miRNAs are connected to the cell cycle regulation, particularly regulation of mitosis.

4.3.2 Hub miRNAs regulate genes enriched in inflammatory pathways

In addition, by counting the minimum and maximum number of targets for the DE miRNAs, we found as few as 9 and as many as 740 genes being targeted by the same miRNA. We hypothesized that miRNAs targeting a high number of genes with an altered expression between M^{un} and $M^{LPS+IFNγ}$ might play important roles in macrophage polarization. Thus, we identified 9 miRNAs that target more than 500 DE genes and we termed these as "hub miRNAs" (Figure 4.5, supplemantry table 4 https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantary_data4.xlsx). Out of these, 5 were upregulated and 4 were downregulated (**Table 4.1**)., but all 9 hub

miRNAs were predicted to bind 79% of all the DE genes targeted by any differentially expressed miRNA, reinforcing their importance.

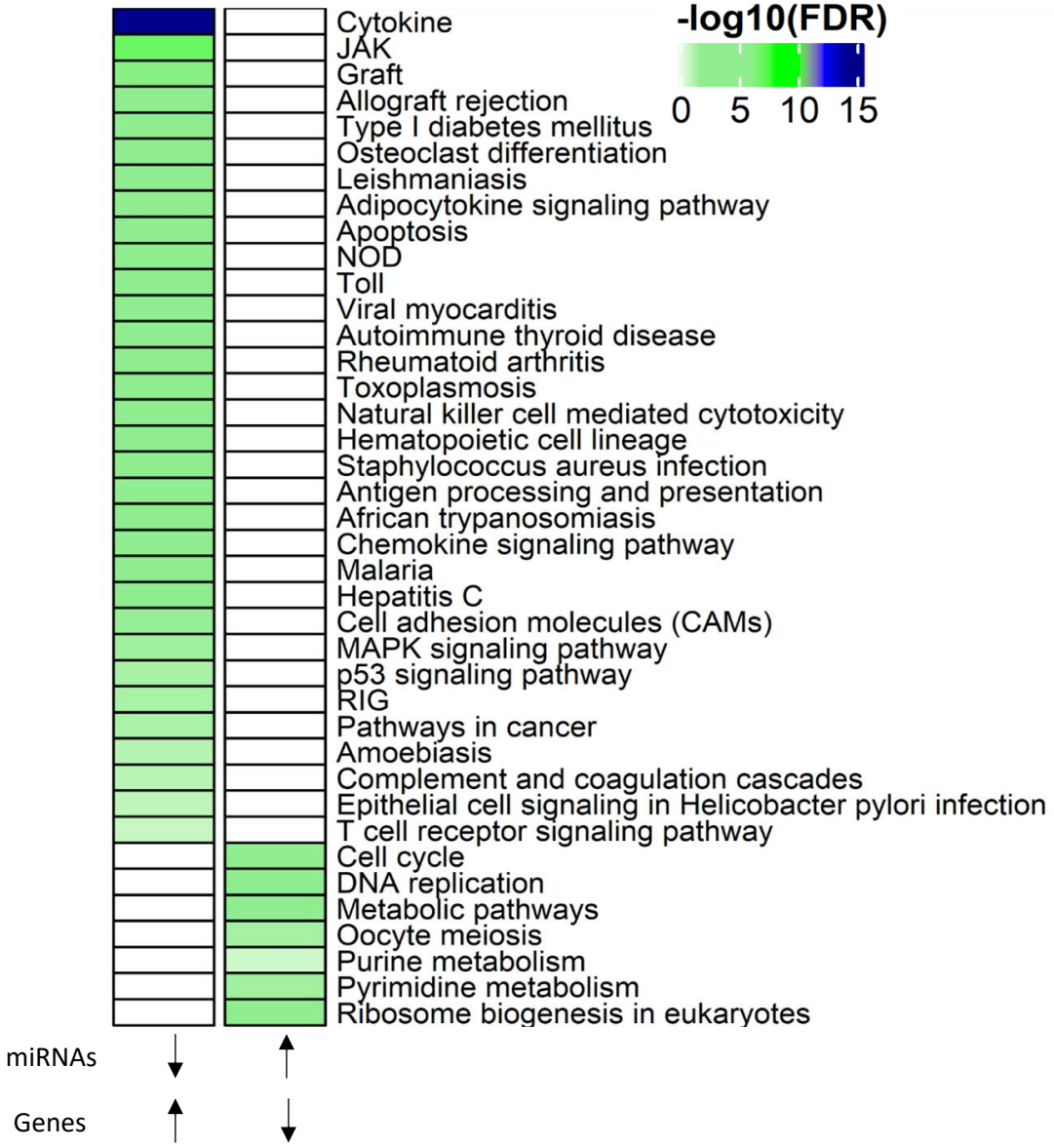


Figure 4. 3- Highly enriched KEGG enrichment pathway analysis of genes differentially expressed between M0 and M1 macrophages (FDR<0.05)

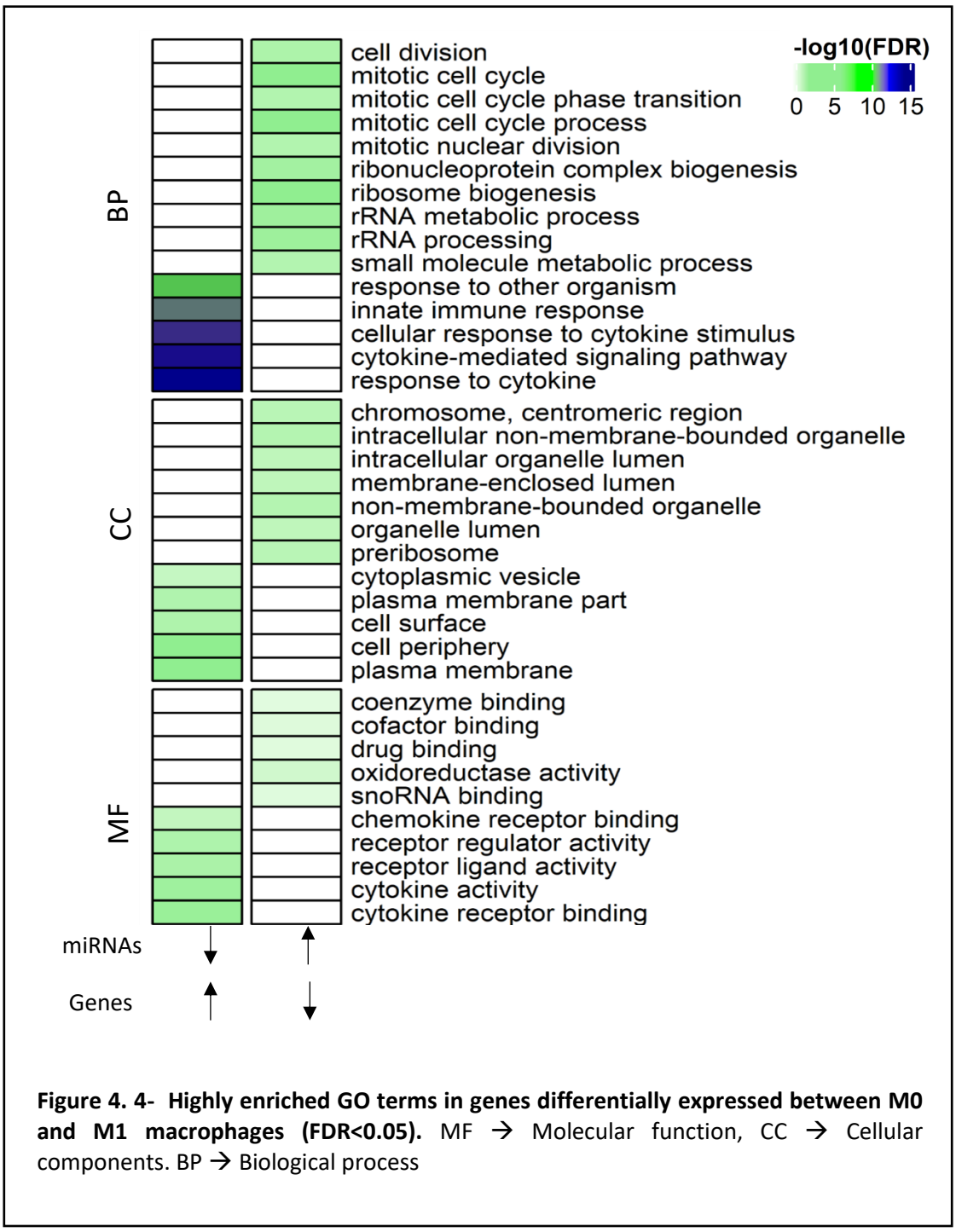


Figure 4. 4- Highly enriched GO terms in genes differentially expressed between M0 and M1 macrophages (FDR<0.05). MF → Molecular function, CC → Cellular components. BP → Biological process

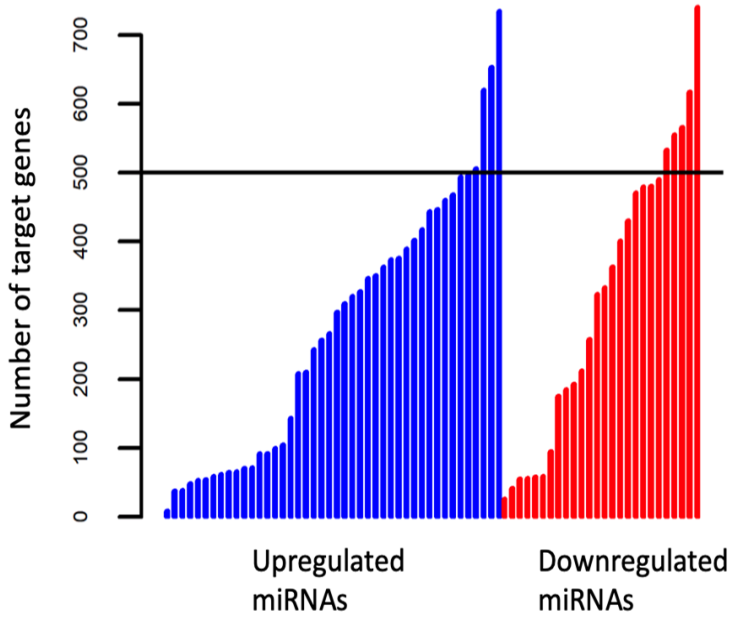


Figure 4. 5- Number of differentially expressed genes, potentially regulated by differentially expressed miRNAs (M0 vs M1)

Table 4. 1- The number of target genes of 9 "hub miRNAs" and the positive control miR-155-5p, shown in the last row

miRNA	Expression in M1 vs. M0	No. of target genes
miR-186-5p	Upregulated	740
miR-4773	Upregulated	617
miR-3614	Upregulated	566
miR-125a	Upregulated	555
miR-7-5p	Upregulated	533
miR-335-3p	Downregulated	734
miR-766-3p	Downregulated	653
miR-1343-3p	Downregulated	620
miR-4709-3p	Downregulated	506
miR-155-5p	Upregulated	363

Gene enrichment analysis of target genes for both downregulated and upregulated hub miRNAs proved successful in identifying potential biological functions of both of these classes of miRNAs. Thus, we observed that the ten most significantly enriched pathways amongst downregulated targets for each upregulated miRNA were involved in chromosome assembly and condensation factors (**Figure 4.6**). Similarly, most upregulated targets of downregulated hub miRNAs had roles in cytokine, T cell and INF signaling pathways alike (**Figure 4.7**).

We further tested our prediction by looking at miR-155-5p as a positive control for our analysis, which is an miRNA conserved across vertebrates and known to regulate inflammatory responses in macrophages (Curtale *et al.* 2019). As expected, we found it to be significantly upregulated in $M^{LPS+INF\gamma}$ and predicted it to bind 363 downregulated targets. While enrichment of "DNA-dependent DNA replication" amongst targets was not significant for miR-155-5p, a non-significant 3.1 fold enrichment was present. We believe the lack of significance is probably due to the smaller total set of targets.

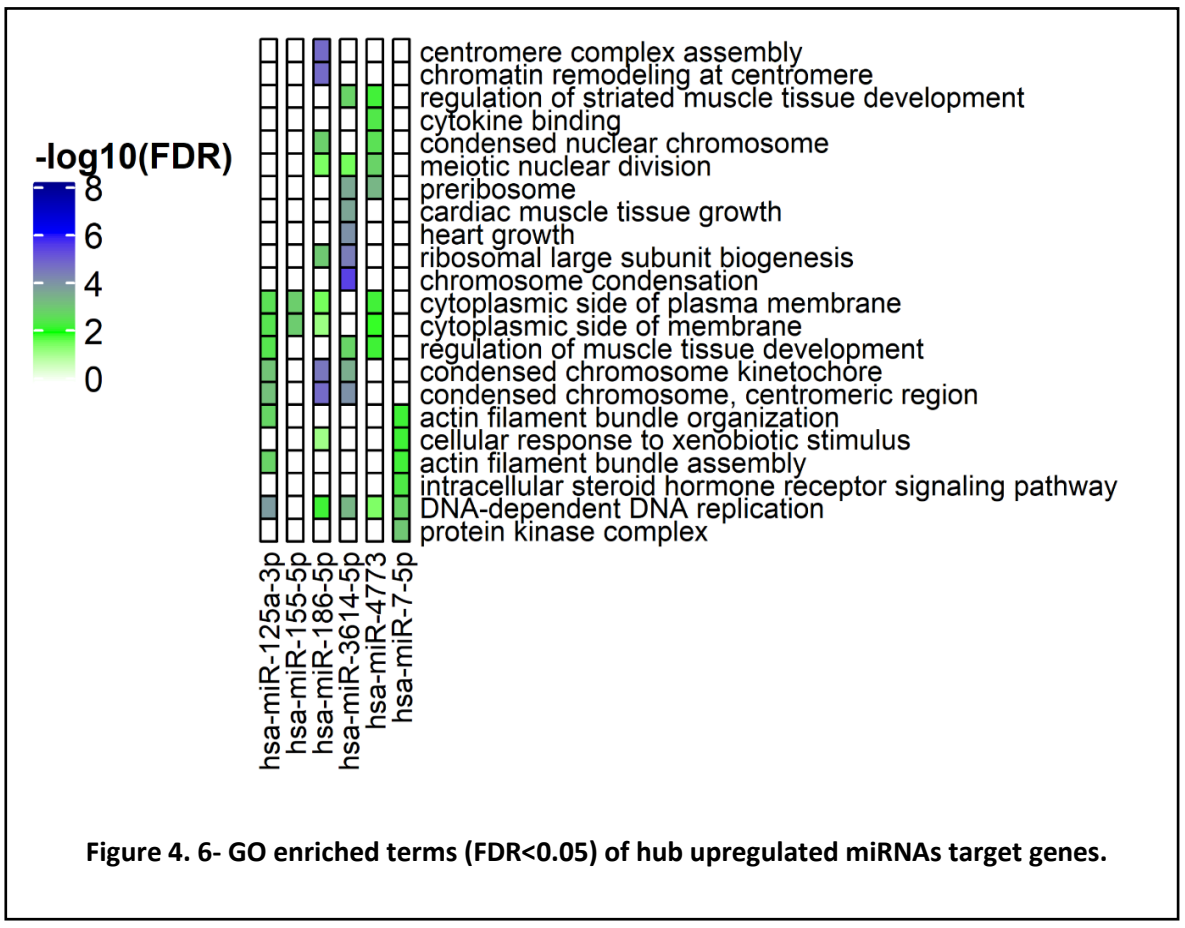


Figure 4. 6- GO enriched terms (FDR<0.05) of hub upregulated miRNAs target genes.

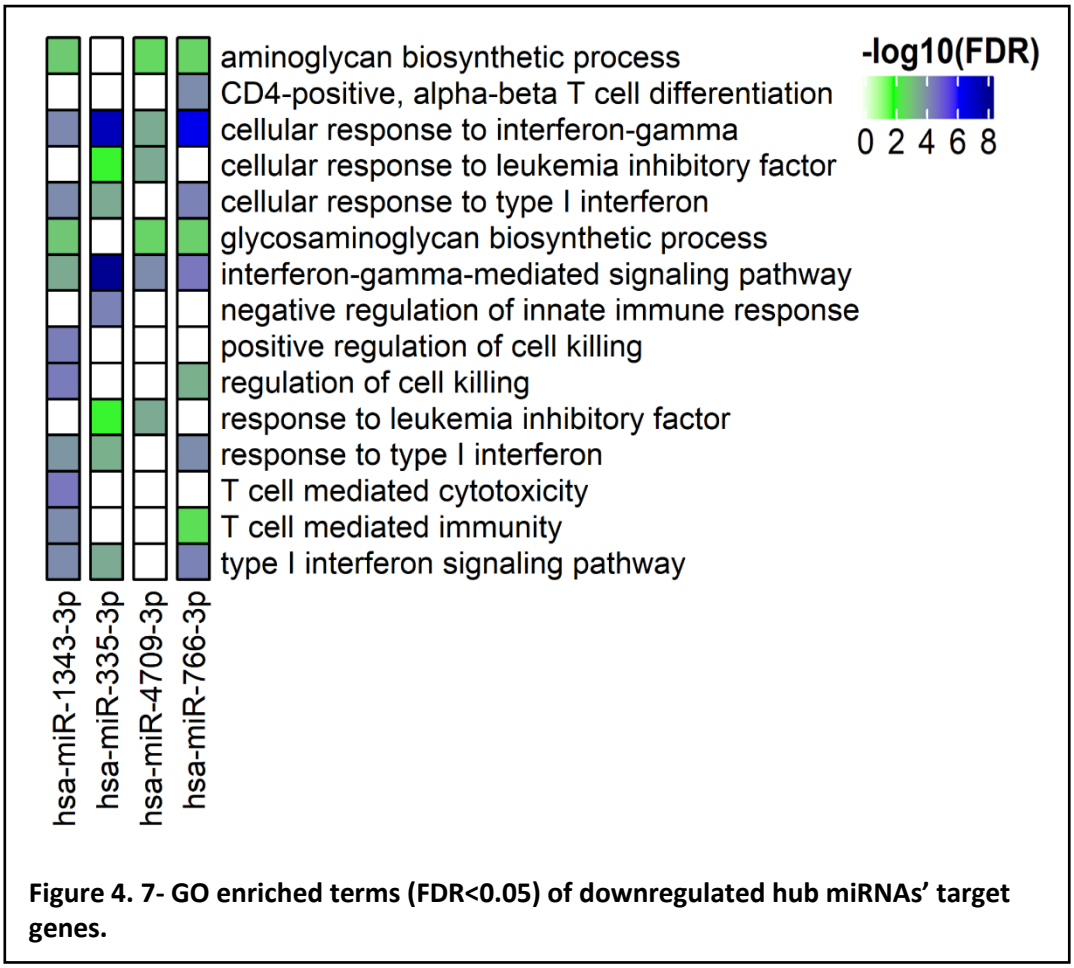


Figure 4. 7- GO enriched terms (FDR<0.05) of downregulated hub miRNAs' target genes.

We also observed that upregulated target genes involved in INF signaling pathways were shared by 3 of the 4 downregulated hub miRNAs. Upon a more in-depth analysis, we found that downregulated hub miRNAs share approximately 30-35% of their targets **(Figure 4.8**). In contrast, upregulated miRNAs shared overall fewer targets, between 20-25% of them (**Figure 4.9**). Similarly, looking at GO enrichment profiles, we saw that upregulated genes of downregulated hub miRNAs share more than downregulated genes of upregulated hub miRNAs, but they all have distinct GO profiles, and only "DNA-dependent DNA replication" being common across all 5 hub miRNAs.

Outcomes of these analyses were consistent with our previous pathway analysis for the putative targets of DE miRNAs, suggesting that the hub miRNAs may play an important role in shaping the inflammatory response and polarisation of macrophages.

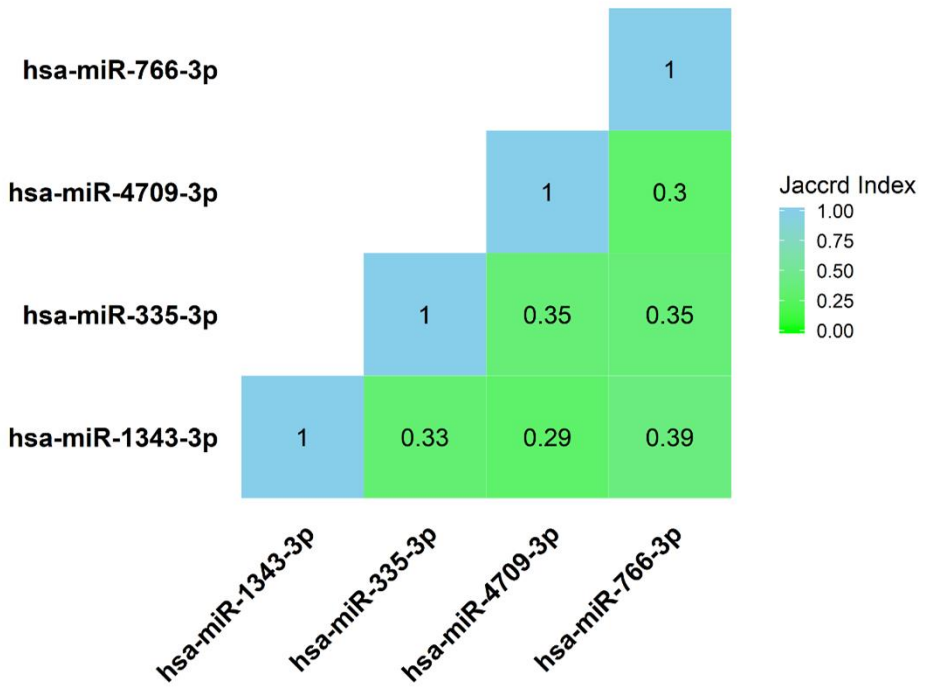


Figure 4. 9- Common target genes between hub downregulated miRNAs, showed in pairs and calculated using the Jaccard Index.

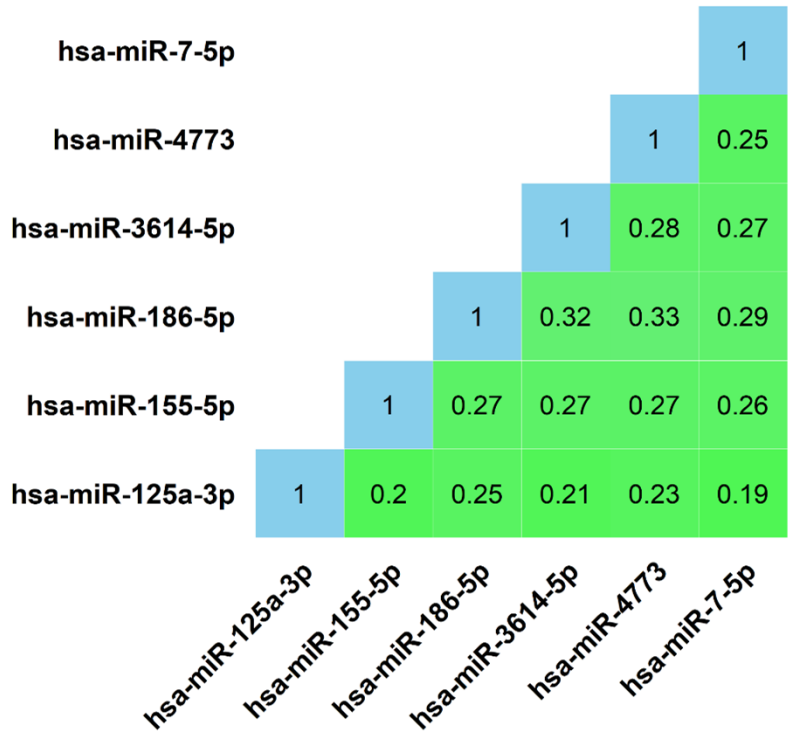


Figure 4. 8- Common target genes between hub upregulated miRNAs, showed in pairs and calculated using the Jaccard Index.

4.3.3 Three hub miRNAs were confirmed to target genes involved in cell-cycle regulatory pathways

As described above, through our computational analyses we observed that a number of hub miRNAs were significantly up and downregulated in M1 compared to M0 macrophages. We further validated these results, by testing the expression of 3 hub miRNAs through RT-qPCR, miR-125a-3p, miR-186-5p and miR-155-5p considered as control. Our results showed that miR-155-5p and miR125a-3p were significantly upregulated in $M^{LPS+INF\gamma}$ compared to M^{UN} samples (p-value < 0.0001 and p = 0.003 respectively), but not in IL-4 treated macrophages (p-value = 0.5 and p =0.9 respectively). The expression of miR-186-5p was also high in $M^{LPS+INF\gamma}$ cells compared M^{UN}, but it was not at all detected in IL-4 treated macrophages (M2 macrophages) (**Figure 4.10**).

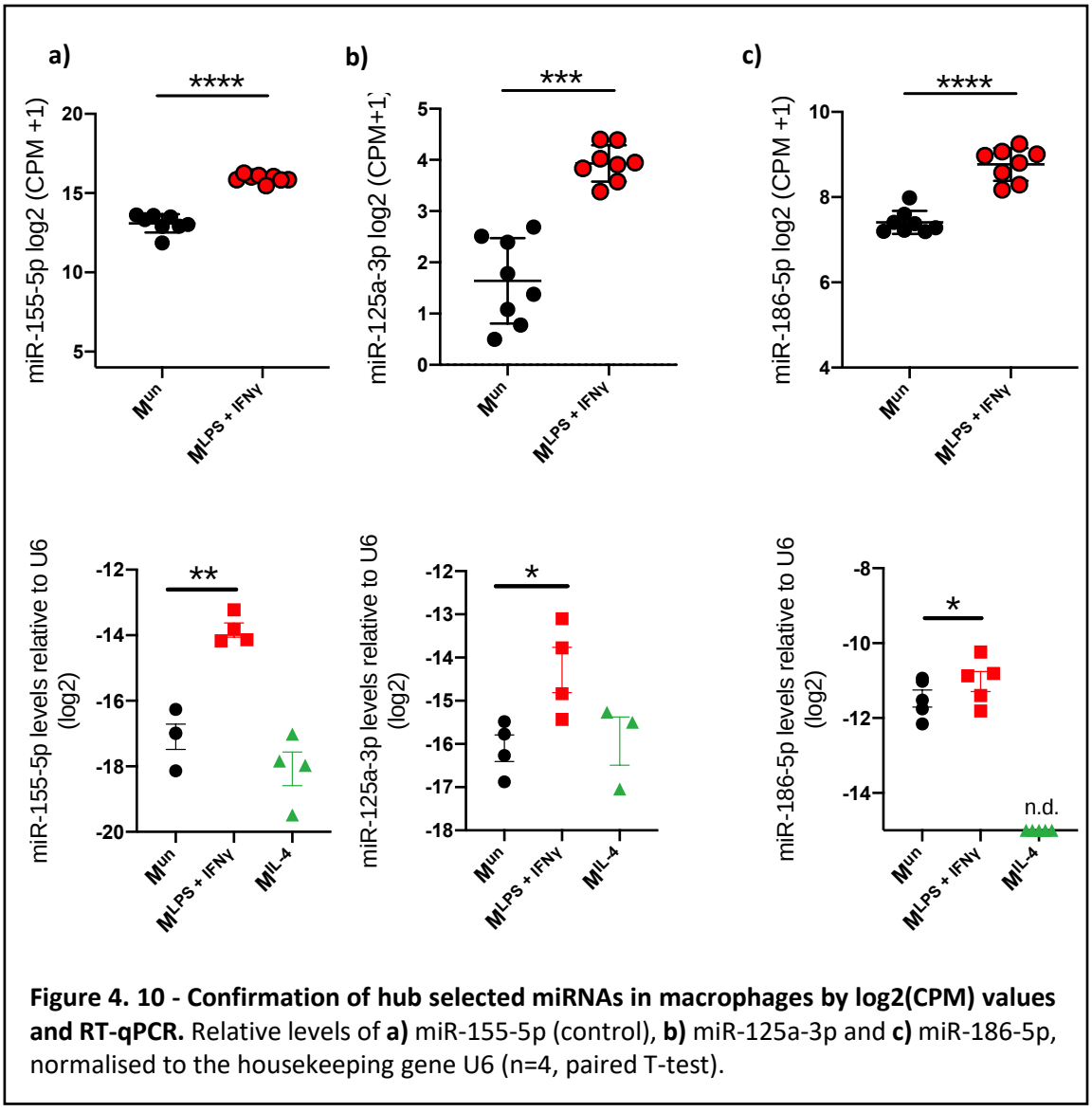


Figure 4. 10 - Confirmation of hub selected miRNAs in macrophages by log2(CPM) values and RT-qPCR. Relative levels of **a)** miR-155-5p (control), **b)** miR-125a-3p and **c)** miR-186-5p, normalised to the housekeeping gene U6 (n=4, paired T-test).

Similarly, RT-qPCR values showed a successful overexpression of the mimics synthesised for all the three hub miRNAs, miR-125a-3p, miR-186-3p and miR-155-5p (control), which were substantiated in M^{UN} samples, in order to further test the role of these hub miRNAs in triggering an inflammatory phenotype in macrophages (**Figure 4.11**).

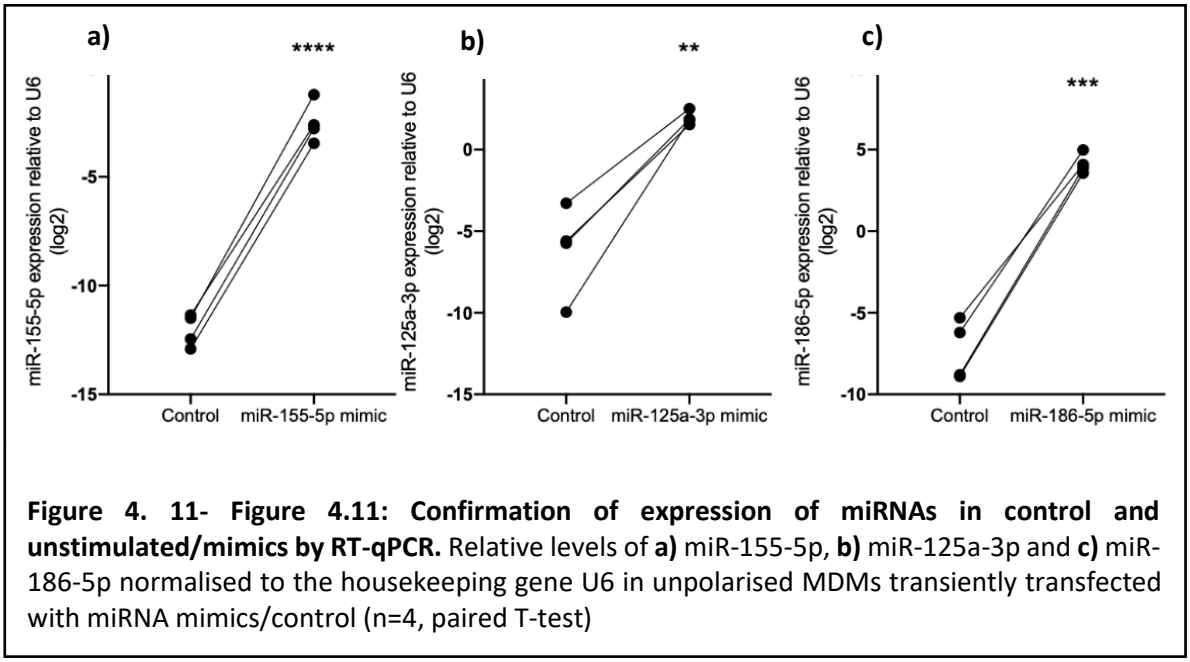


Figure 4. 11- Figure 4.11: Confirmation of expression of miRNAs in control and unstimulated/mimics by RT-qPCR. Relative levels of **a)** miR-155-5p, **b)** miR-125a-3p and **c)** miR-186-5p normalised to the housekeeping gene U6 in unpolarised MDMs transiently transfected with miRNA mimics/control (n=4, paired T-test)

Furthermore, RNAseq analysis of the samples overexpressing the 3 hub miRNAs mimics and subsequent PCA analysis revealed that samples were separated based on self-reported sex (**Figure 4.12**). In fact, response to some miRNAs was consistent between self-reported sexes, e.g. samples overexpessing miR-155-5p were skewed up and to the left of the negative controls for a donor. However, the transcriptomic response to the other mimics was different, e.g. samples overexpressing miR-186-5p appeared below for donors a,c and f, but above for donors b,d,e compared to negative control. Unfortunately, we did not have a sufficient sample size to analyze this effect further.

Looking at the the direction of change for target genes, we observed there were more downregulated than upregulated targets for miR-155-5p and miR-186-5p, upon overexpression of the cognate miRNAs, with 280 of 362 being downregulated for miR-155-5p, and 413 out of 739 for miR-186-5. However, for miR-125a-3p only 219 of 555 targets were downregulated, thus only slightly more targets were upregulated, 336 out of 555. (**Figure 4.13**)

What is more, the targets of both miR-155-5p and mir-186-3p were significantly downregulated compared to non-targets, upon transfection with the mimics ($p<3 \times 10^{-16}$ and, $p<0.004$ for miR-155-5p and miR-186-2p respectively, Wilcoxon test), demonstrating the validity of our predictions, that miRNAs are directly downregulating those target genes. However, our predicted

targets for miR-125a-3p were not downregulated by mimic transfection, which indicates miR-125a-3p may not have a direct impact on these target genes in unstimulated macrophages, but it remains unclear if this is also true in their polarised counterparts. (**Figure 4.14**)

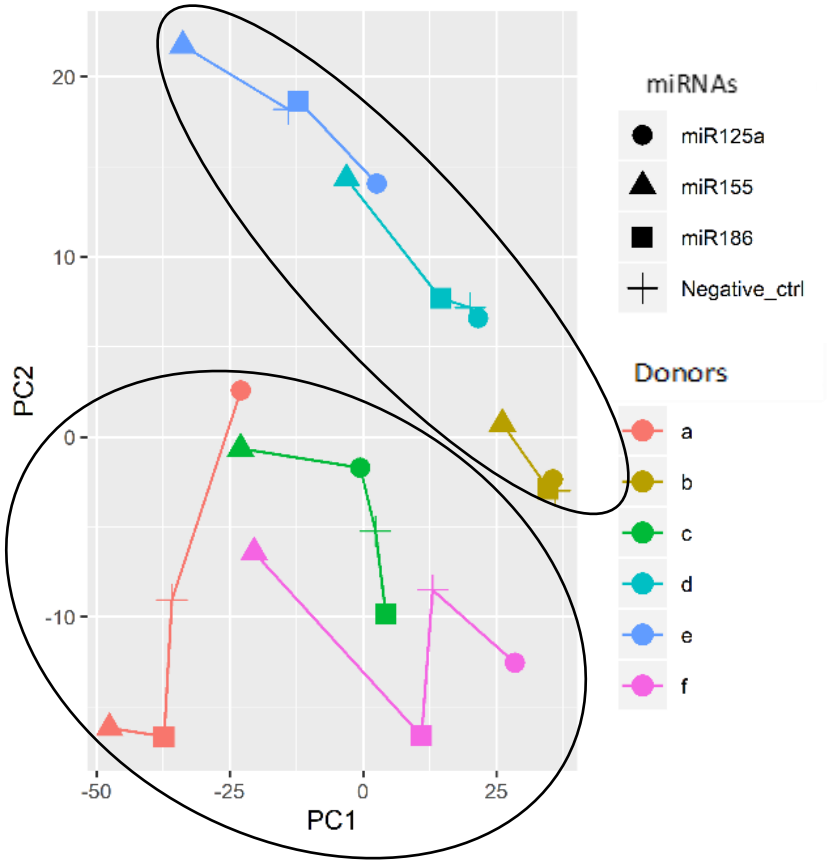


Figure 4. 12- Global PCA plot based on log2 (CPM+1) values from RNA-seq performed on unpolarised MDMs transiently transfected with miRNA performed on unpolarised MDMs transiently transfected with miRNA mimics/control. Circles separates the self-declared sex.

In addition, gene enrichment analysis showed that downregulated targets of miR-155-5p were enriched for genes involved in regulating G2/M transition and DNA replication. In contrast, upregulated targets of this miRNA were involved in B cell activation, proliferation, and chemokine signaling pathway. Downregulated targets of miR-186-5p had roles in chromosome and DNA-packaging, while upregulated ones were linked to positive regulation of lipid kinase activity. Genes downregulated by miR-125a-3p were found to be involved in antigen processing and presentation, while upregulated targets were associated with inositol phosphate metabolism (**Figure 4.15, supplementary table 5,** https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantary_table_5.xlsx)

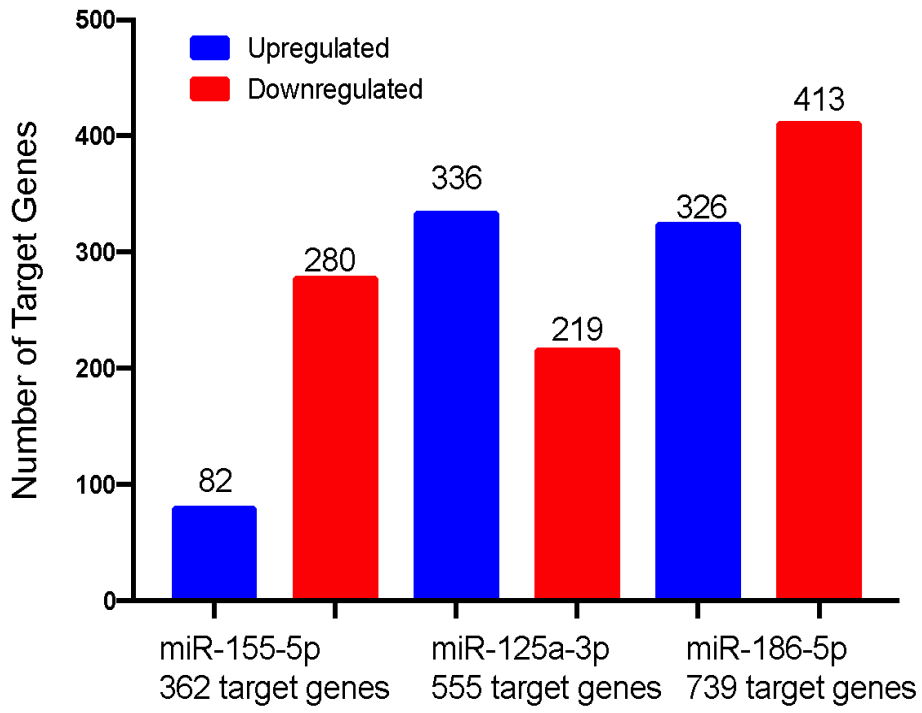


Figure 4. 14- Bar plot showing number of predicted target genes and there regulation

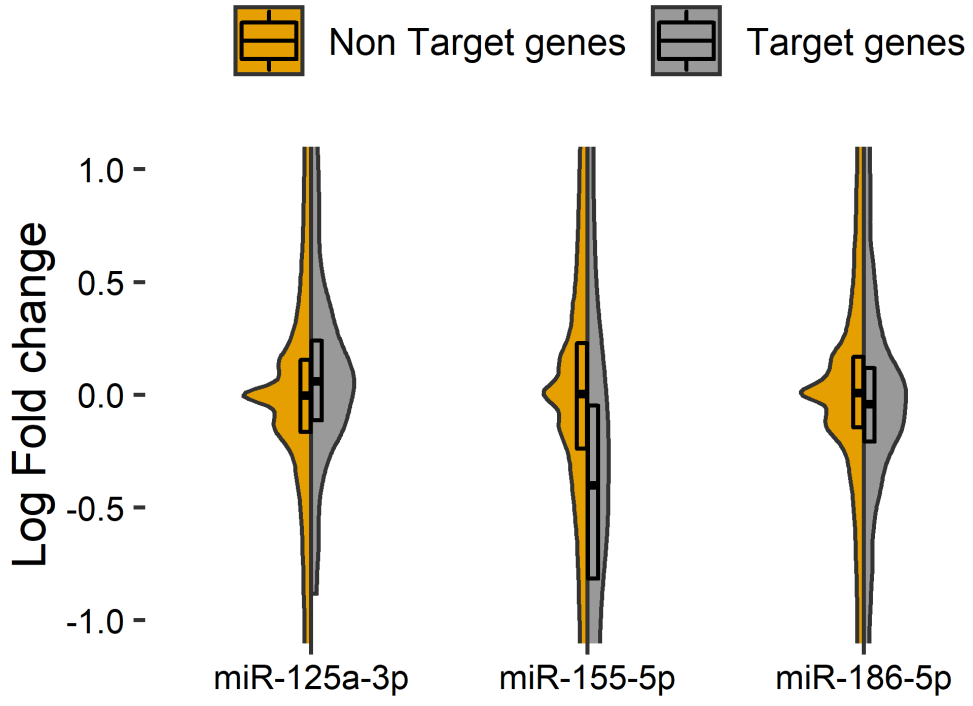


Figure 4. 13- Violin plot showing the change in the expression (log2 Fold Change) level between target and non-target genes of each miRNA, overexpressed in MDMs.

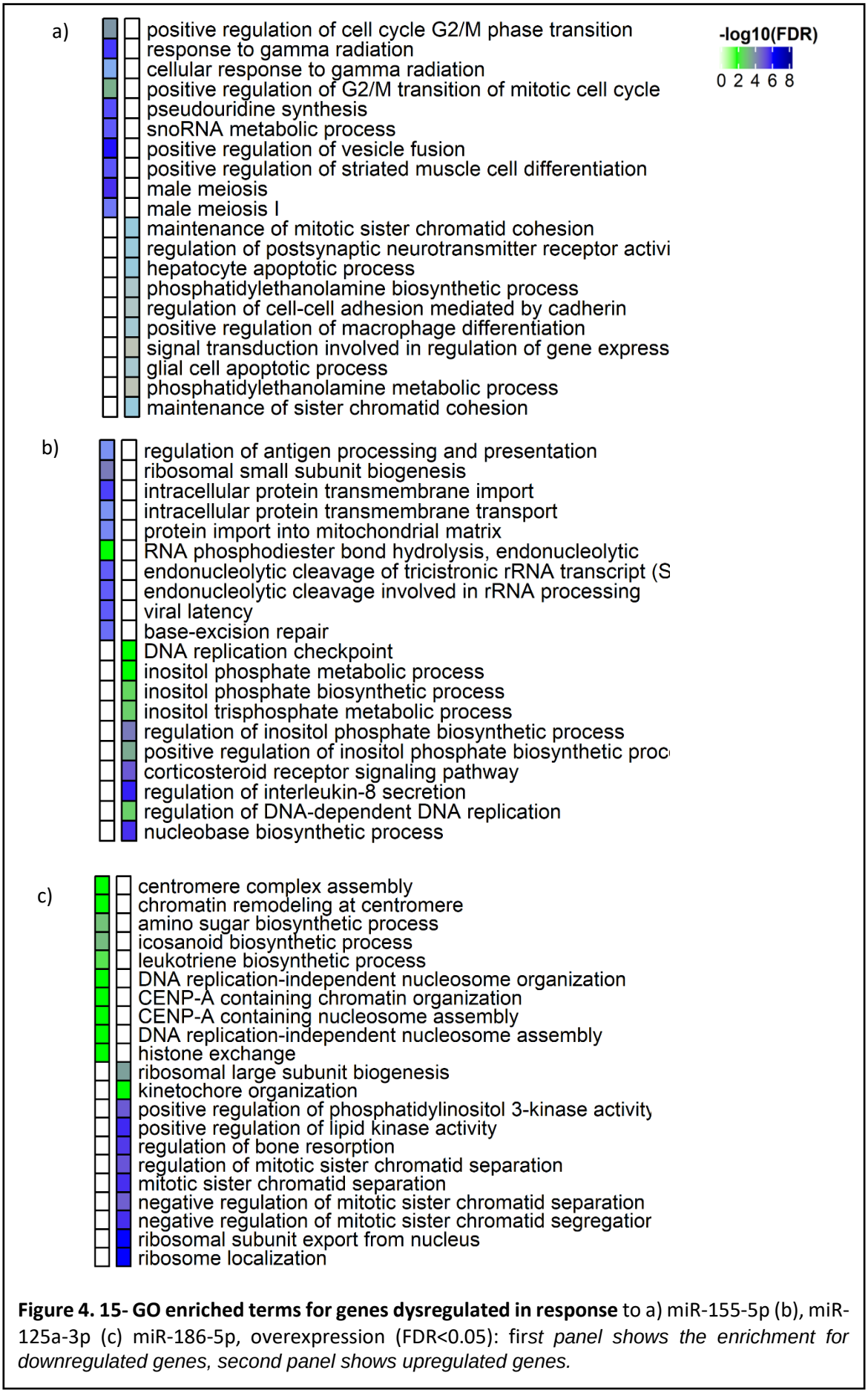


Figure 4. 15- GO enriched terms for genes dysregulated in response to a) miR-155-5p (b), miR-125a-3p (c) miR-186-5p, overexpression (FDR<0.05): fir*st panel shows the enrichment for downregulated genes, second panel shows upregulated genes.*

4.3.4 Loss of miRNA binding sites through APA processing can help regulate gene expression

All the above analyses were performed on 3'UTR sequences of standard length (without APA processing). As mentioned in this chapter's introduction, APA can increase 3'UTR diversity and affect transcript expression by altering miRNA binding sites. Therefore, the following analysis was focused on changes in 3'UTR lengths caused by APA events in both M^{un} and $M^{LPS+INF}$. Overall, 6188 transcripts of 2395 genes showed a change in their 3'UTR lengths, out of which 2661 had their 3'UTRs shortened. 431 of these (16%) were uregulated in $M^{LPS+INF}$ compared to M^{un}, while 206 (7%) were downregulated. The density plot in **Figure 4.16 (**supplemantry table 6, https://github.com/srmeetd/Macrophages_transcriptomic_analysis/blob/main/Supplemantary_data6.xlsx**)** shows the change in percentile of 3'UTR lengths in macrophages, generated by DaPars as sites usage, with the positive number representing shorter 3'UTR sites/site usage in $M^{LPS+INF}$, and negative values showing an increase in 3'UTR sites usage in M^{un}.

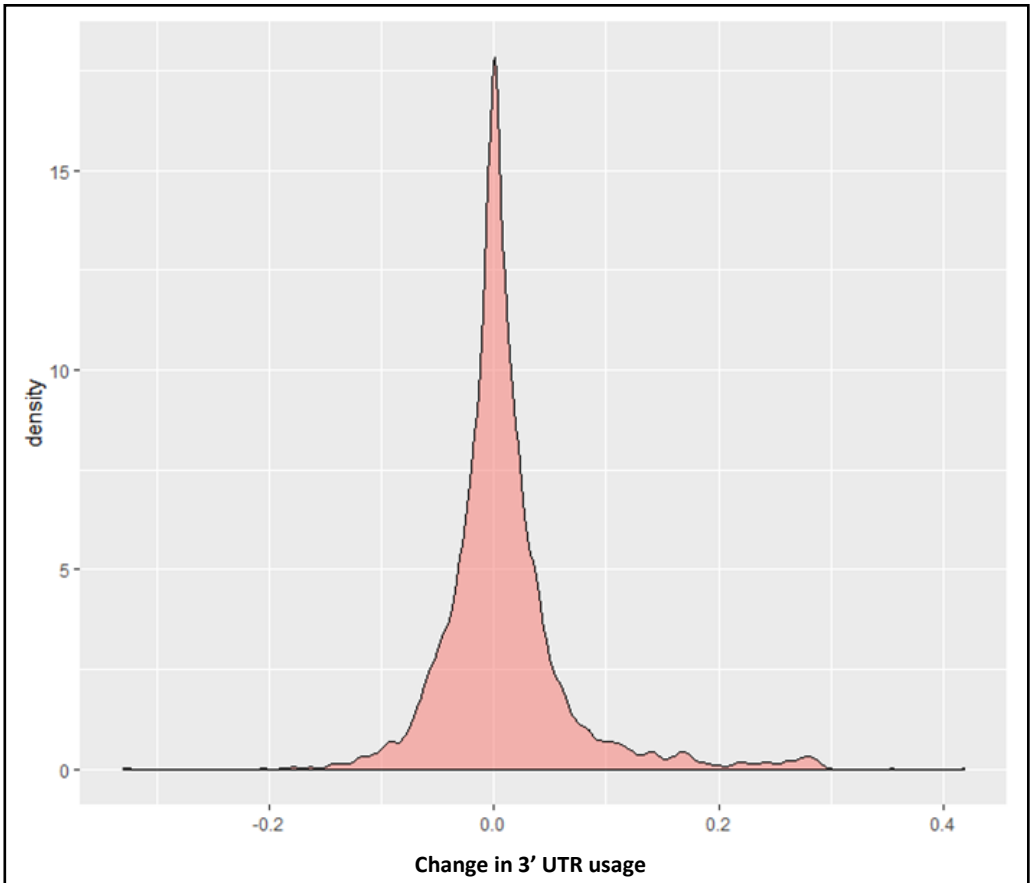


Figure 4. 16- Density plot representing changes in 3'UTR length/site usage in Mun and MLPS+IFG. Positive numbers represent shorter 3'UTR sites/site usage in stimulated cells, and negative values show an increase in 3' UTR sites usage in unstimulated.

In order to understand if miRNAs are involved in macrophage polarisation, we hypothised that when miRNAs are upregulated, their targets would be downregulated and viceversa – downregulated miRNAs would lead to upregulated target genes. Moreover, we also think that upregulated miRNAs whose binding sites are degraded would also lead to an upregulation in their target genes. One way miRNA binding sites could be removed would be through 3'UTR shortening *via* APA processing.

In this sense, we observed that binding sites for 17 out of 26 upregulated miRNAs were lost in a total of 108 out of 1460 up-regulated genes in $M^{LPS+INF}$ due to APA events, with 20% of these 108 upregulated target genes being shared between the 17 miRNAs (**Figure 4.17**). In addition, out of 17 miRNAs, 5 miRNAs were found to be hub-miRNAs, responsible for downregulating genes involved in cell-cycle-related pathways. This result supports our hypothesis and, considering the majority of these target genes are involved in cell signalling and inflammatory-related pathways specific to an M1 phenotype (**Figure 4.18**), their 17 miRNAs may also play an important role in macrophage polarization. More than that, genes involved in signaling pathways important in polarizing M1-like macrophages may, in fact escape post-transcriptional regulation by shortening their 3'UTRs.

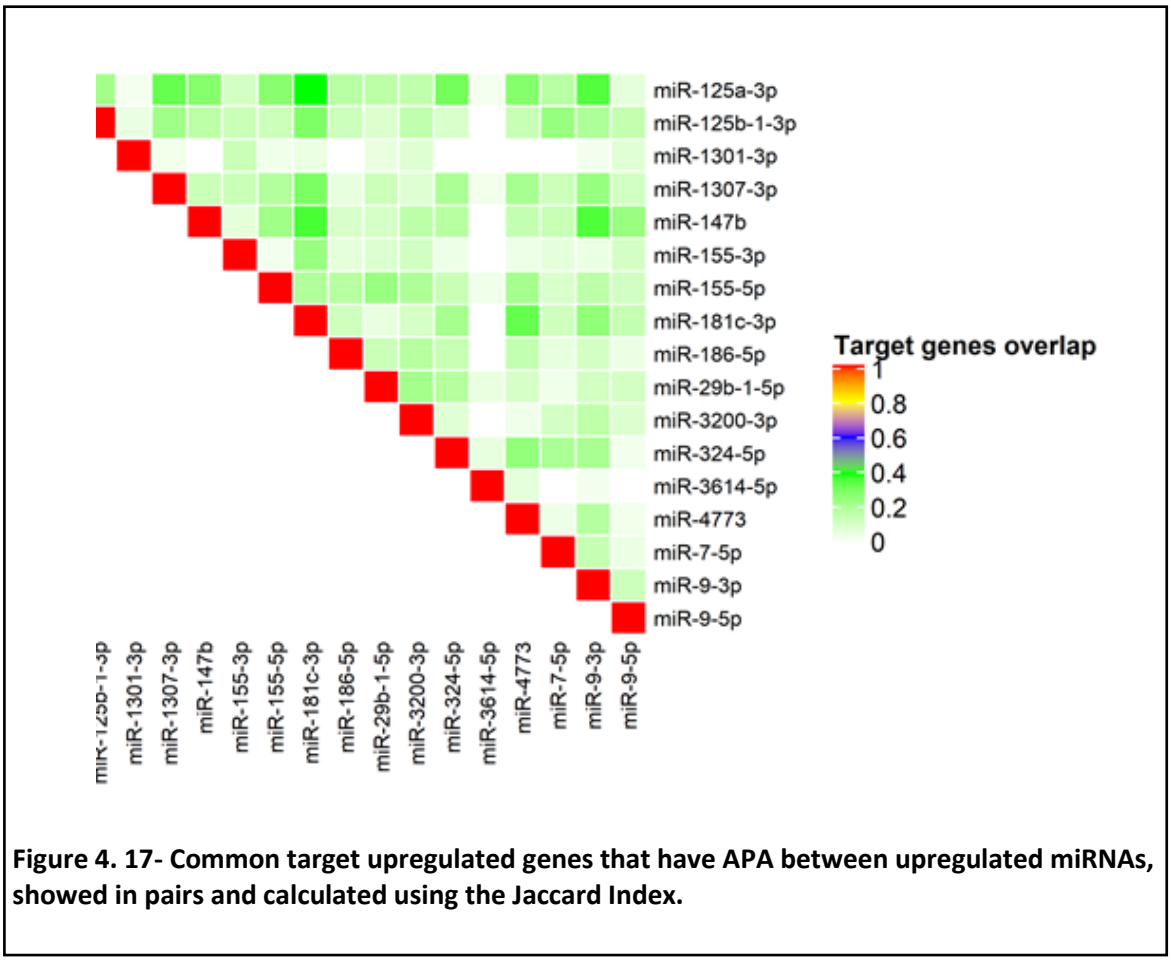


Figure 4. 17- Common target upregulated genes that have APA between upregulated miRNAs, showed in pairs and calculated using the Jaccard Index.

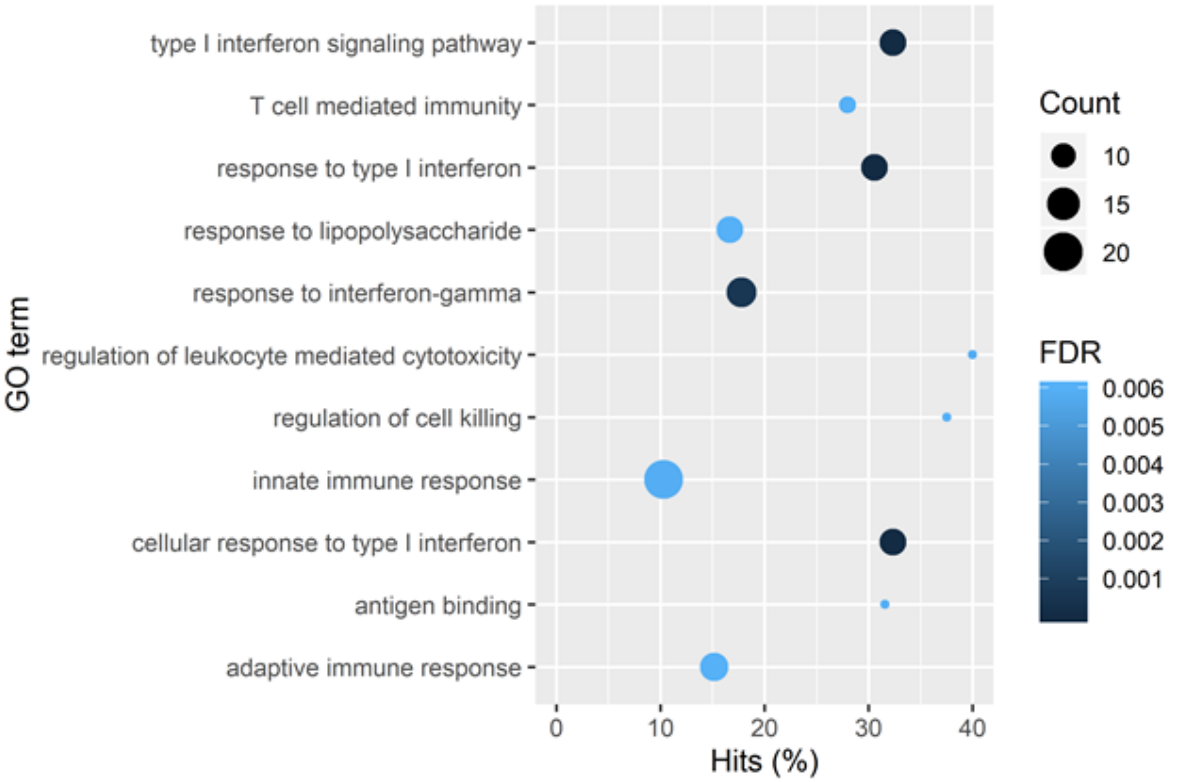


Figure 4. 18- GO enriched terms for upregulated genes in MLPS+IFNγ lost binding sites of upregulated miRNAs due the APA process. Dot plot represents top 10 gene ontology annotations, x-axis shows the percentage of differentially expressed genes present in particular ontology function, y-axis represents gene ontology annotations, dot size represents the number of differentially expressed present in corresponding gene onotology function and color represents the FDRscore.

As described above, in M1 phenotype, loss of binding sites (*via* APA processing) of upregulated miRNAs can lead to upregulation of their target genes. In order to test whether the same can be observed in M0 macrophages, we looked at miR-125a-3p as one of the five hub miRNAs identified (Chapter 4.3.3). Compared to two other miRNAs, also tested *in vitro* (miR-155-5p and miR-186-5p), miR-125a-3p was the only one whose target genes were not downregulated compared to its nontarget genes, in unstimulated cells.

Moreover, its putative target genes were observed to not be affected by this miRNA in unstimulated macrophages transfected with a miR-125a-3p mimic (**Figure 4.14**). We believe the reason for this could be that its target genes were lacking the binding sites for miR-125a-3p in the unstimulated samples. To test this hypothesis, we looked at the changes in length of 3'UTRs of miR-125a-3p putative target genes (**Figure 4.19**) that rose from degradation of miRNA binding sites upon APA processing, including those of miR-125a-3p. In total, we found that out of 555 downregulated

genes, 35 presented APA events, but only 14 of these did not have binding sites for miR-125a-3p. Our hypothesis was thus rejected, as the results suggested that miR-125a-3p has, in fact, little to no impact on negatively regulating the expression of its putative target genes in M0 condition.

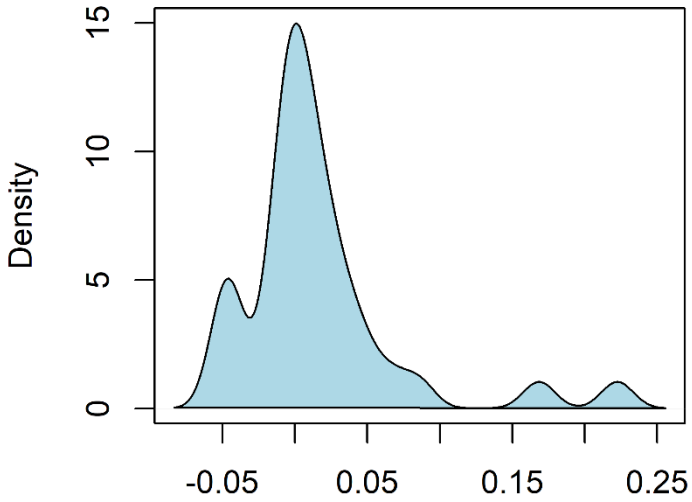


Figure 4. 19- Density plot representing change in 3' UTR usage of a genes Mun and MLPS+IFG. Positive number represents less UTR usage in $M^{LPS+IFG}$ samples and negative sites represents less UTR usage in M^{un} samples

4.4 Discussion

In this chapter, we described the role of miRNAs in macrophage polarization, particularly in M0 and M1 phenotypes, the latter being specific to pro-inflammatory responses. What we found was that **a)** Differentially expressed miRNAs preferentially target genes involved in pro-inflammatory pathways and that **b)** Hub miRNAs regulate genes enriched in inflammatory pathways. **c)** Three hub miRNAs were also confirmed to target genes involved in cell-cycle regulatory pathways and **d)** Loss of miRNA binding sites through APA processing was shown to help regulate gene expression.

Differential expression analysis of miRNAs between M^{un} and $M^{LPS+IGNg}$ showed a clear separation between samples in the two conditions, which strengthens the belief that miRNAs are important factors for macrophage differentiation. However, this result is yet insufficient to understand specifc roles of miRNAs in cell polarisation. Therefore, by also selecting only differentially expressed target genes of miRNAs (**Figure 4.2**), we found 1767 downregulated targets of 26 upregulated miRNAs, and 1531 upregulated targets of 44 downregulated miRNAs. What is more, gene enrichment analysis on these targets (**Figures 4.3** and **4.4**) showed that upregulated genes are mainly involved in cytokines-related pathways, such as JAK/STAT, Toll-like receptor, chemokine signaling

pathways and NOD-like receptor pathways. This result further confirms the involvement of miRNAs in polarising unstimulated macrophages towards an M1 phenotype, and offers a better insight into what particular processes they are most likely to regulate transcriptionally.

On the other hand, gene enrichment analysis on downregulated targets of upregulated miRNAs showed these genes are mostly responsible for activating cell-cycle pathways. However, their expression being downregulated by miRNAs suggest the cell-cycle pathways they are involved in may actually be arrested, which are often important for controlling inflammation upon pathogen infections. All in all, these results imply that one miRNA has multiple targets, which raises another question – are all miRNAs important in macrophage polarisation, or is there just a set of miRNAs targeting genes that play key roles in polarising macrophages?

In this sense, we counted the number of targets for each miRNA and identified a set of 9 miRNAs binding 79% of all the DE genes, which we named "hub-miRNAs". Out of these 9 hub miRNAs, 5 were upregulated and 4 were downregulated (**Figure 4.5, Table 4.1**), and gene ontology enrichments revealed that targets of all the 9 hub miRNAs are mostly involved in inflammatory pathways and cell cycle-related pathways, being a subset of pathway enrichment shown in **Figures 4.6** and **4.7**. Moreover, we observed that approximately 30-35% of upregulated target genes are shared between the 4 downregulated hub miRNAs and 20-25% of downregulated targets shared by the 5 upregulated miRNAs (**Figure 4.8 and 4.9**).

Therefore, our results described above show that not all differentially expressed miRNAs are important in macrophage polarisation, but the network of 9 hub miRNAs and their targets do play an important role in this process. Future work could focus only on hub miRNAs, in order to understand the mechanism of macrophage plasticity, and it would be interesting to see the effect of such macrophage plasticity on the knockdown of upregulated miRNAs in $M^{LPS+IFN\gamma}$ phenotype.

What is more, these results were based on putative target genes obtained from the TargetScan database, and in order to further validate our computational observations, we selected three upregulated miRNAs and overexpressed them in M^{un} macrophages, one of which being the hub miRNA miR-155-3p, well known for its role in macrophage polarisation. *In vitro* validation of these three miRNAs confirmed that their target genes were more downregulated compared to non-target genes, except those of miR-125a-3p (**Figure 4.14).** A possible reason for this could be that this miRNA may have low to no impact on the expression of its targets, or that target genes may in fact skip post-transcriptional regulation by undergoing APA processing, which can remove the binding site for miR-125a-3p through 3'UTR shortening.

Moreover, by showing that the downregulated targets of these three miRNAs are involved in cell-cycle-related pathways (**Figure 4.15**), we confirmed that our methodology used for an integrated miRNA-mRNA analysis and previous interpretations were accurate. However, all the above analysis was performed on standard 3'UTRs of genes recorded in TargerScan database. Therefore, we further looked at the effect of 3'UTR shortening through APA processing on miRNA binding sites, and we observed that 3'UTRs of DE genes in $M^{LPS+IFN\gamma}$ were indeed shorter than in M^{un}. More of these target genes were found to be required in immunity-related pathways and our results show that they may undergo APA events in order to shorten their UTRs and avoid the binding of RNA-binding proteins or miRNAs that could downregulate their expression.

In total, we identified 17 upregulated miRNAs binding sites being removed by APA processes across 108 upregulated genes in $M^{LPS+IFN\gamma}$. What is more, KEGG pathway analysis confirmed that these 108 genes were involved in both cell signaling and inflammatory pathways (**Figure 4.18**). All in all, the above results also atested that APA processing can be important in regulating the expression of genes important in immunity-related pathways. However, only 14 out of 555 downregulated targets of miR-125a-3p were found to undergo APA processing in M^{un} samples. This hints that miR-125a-3p has less or no impact on the expression of its targets, but this could be an isolated example observed, and in order to formulate a clear conclusion, further analysis could be performed on larger sample sizes.

Chapter 5

Conclusions and Future Work

TRIB1 has been shown to have various biological functions, for example in cell polarization, migration and signalling pathways, and it has also been associated with a number of affections, such as different types of cancer, atherosclerosis, diabetes and cardiovascular diseases (Johnston et al. 2019; Kathiresan et al. 2008; Liang et al. 2013; Niespolo et al. 2020). Despite its therapeutic potential, TRIB1's variation in expression levels across different tissue types continues to pose a challenge for understanding its genomic network. This variation in expression has been observed at both gene and protein levels, which were often found not to correlate with each other and poses another challenge in identifying the exact requirement for TRIB1 protein within the cells (Soubeyrand et al. 2016). Because of these limitations, the co-expression and interaction network of Trib1 have not be fully characterised and so far, studies have only validated TRIB1's relationship with CEBPα and COP1.

For the scope of this project, we first described the importance of miRNAs in macrophage polarization and the role of SNPs altering miRNA binding sites in both Trib1, as well as across the whole genome in allele-specific expression (ASE). In this sense we developed and validated a new pipeline for detecting ASE variants, without prior genotypic information. In addition, we have further investigated Trib1's expression, regulation and function, and we presented our findings on the genomic network of TRIB1 in cancer and macrophages. Thus, we observed that variations in Trib1's 3'UTR, as well as in the miRNA binding sites are directly associated with changes in its expression levels. We also reported Trib1 co-expresses with early response genes and we proposed a novel hypothesis of Trib1 potentially being an early response gene, even acting as an E2 ligase.

○ Role of miRNAs in macrophage polarisation

When studying the transcriptional regulation during macrophage polarisation, we found 73 differentially expressed miRNAs between Mun and MLPS+IFNγ targeting 79% of the differentially expressed genes between the two conditions. Out of these, we identified 9 hub miRNAs (targeting more than 500 genes) – 4 downregulated (miR-335-3p, miR-766-3p, miR-1343-3p, miR-4709-3p) and 5 upregulated (miR-186-5p, miR-4773, miR-3614, miR-125a, miR-7-5p). KEGG pathways revealed upregulated targets of downregulated miRNAs are involved in pro-inflammatory-related pathways, while downregulated genes of upregulated miRNAs were linked to cell-cycle progression. Furthermore, we found that 17 upregulated miRNAs lost their binding sites on 108 upregulated targets due to alternative polyadenylation (APA). These genes are involved in cell-signalling and Chapter 5:

Conclusions and Future Work 137 inflammatory pathways, and our results suggest that APA might be an important step in regulating genes of immunity-related processes. In order to confirm our computational results, 3 miRNAs previously identified as upregulated in MLPS+IFNγ (miR-155-5p, miR-186-5p and miR-125a-3p) were validated in vitro, being overexpressed in Mun. RNAseq analysis of these samples showed that target genes of miR-155-5p and miR-186-5p were significantly downregulated compared to non-target genes, but those of miR-125a-3p were not. What is more, only 14 out of the 555 downregulated targets of miR-125a-3p were found to have undergone APA processing, which ruled this out as the possible reason for its targets not being significantly affected compared to the overall non-target gene pool. Therefore, we concluded that miR-125a-3p may have low to no impact on the expression of its targets, or that its target genes may in fact skip post-transcriptional regulation through other means or target genes might not be expressed in MLPS+IFNγ . Future work would require more in vitro testing of the miRNAs identified through RNAseq analyses, in order to validate the efficiency of our computational tools and confirm our observations. These should be performed in both Mun and MLPS+IFNγ , in order to understand the biological effect of these miRNAs on inflammatory-related target genes. Moreover, experiments could be replicated in other cell types, to investigate whether regulation through these miRNAs is universal or tissue specific, and to identify those linked to particular types of cells or processes, such as differentiation. RNAseq would need to be analysed for all these conditions, as well, to compare results with our previous ones, to further test our tools and ensure reproducibility and consistency.

○ The role of SNPs altering miRNA binding sites in ASE

When investigating the post-transcriptional regulation of TRIB1, we identified 60 SNPsderived variants in its 3'UTR, and we observed that the gene expression levels of TRIB1 increased with the number of 3'UTR transcript variants. Upon closer look, miRNA binding sites were altered by these SNPs, with more binding sites being created than degraded in both Mun and MLPS+IFNγ . This suggests that TRIB1's expression could also be regulated through variants with altered binding sites of other mRNA regulatory elements, such as RNA-binding proteins, translational repressors, splicing factors and riboswitches. However, if these variations were affecting TRIB1 expression, we would expect their presence to lead to ASE. Our study of the 60 SNPs detected in the 3'UTR of TRIB1 showed, however, that no significant SNPs in this region were linked to ASE events, but a number of these were associated with disease traits through the Genome-Wide Association Studies (GWAS). This finding suggests that there is link between TRIB1 and different disease trait especially cardio-vascular diseases and auto-immune disease. However, direct link between TRIB1 and disease trait still needs to be explored. Chapter 5: Conclusions and Future Work 138 Although ASE was not detected for any TRIB1 variants, we further tested for SNPs in the 3'UTR of genes across the entire genome that could be

linked to ASE. In this sense, we developed our own pipeline for identifying ASE at both SNP and gene levels, which we validated against different RNAseq datasets. Thus, we observed that SNPs linked to allele specific expression are also responsible in altering miRNA binding sites. More than that, our ASE pipeline helped identify ASE genes in M1 macrophages that were enriched in inflammatory-related pathways. Overall, we observed that a minimum of 1 SNP could be responsible for the allelic imbalance of a given gene, and that genes with the maximum number of significant SNPs identified through our pipeline have already been linked to ASE. This demonstrates the high accuracy of our pipeline (e.g. for TRIB1 we found 90% of SNPs previously annotated) and suggests that in order to maximise the chance of finding true ASE events, both SNPs and genes associated with ASE should be considered. Therefore, we consider our pipeline to be the first complete framework to detect allelic imbalance at both gene and SNP levels, as well as identify altered miRNA binding sites, all without the requirement of prior genotypic information. Future developments could be sought to increase the robustness of our pipeline, for example through validation against different sequencing data types (e.g. DNAseq, ChIPseq). In addition, improvements could be made to reduce the computational power and time requirements, as currently, processing 170 samples in parallel, on an HPC server takes approximately 16 hours. Furthermore, alongside our Readbacked-phasing algorithm, several others could be incorporated as additional options, such as HapTree-x and phASER.

○ Co-expression analysis of TRIB1

In order to better understand the biological function of Trib1, we investigated its genomic network in both macrophages and cancer cell lines, and we found that TRIB1 co-expresses with completely different sets of genes in these different conditions. This was also observed even for different types of macrophages, which could also be linked to variations in gene level expression of TRIB1 between control MDMs, Salmonella typhimurium and Listeria monocytogenes-infected samples. A more in-depth analysis of the co-expression network of TRIB1 in Listeria monocytogenesinfected samples, representative of an M2 phenotype, indicated a potential dual role of Trib1 protein. Firstly, TRIB1 was found to co-express with E3 ligase UNKL and to have a 76% overlap with the structural domain of UBE2D3 E2 ligase. Considering that Trib1 has been previously shown to recruit E3 ligase COP1 for the ubiquitination of CEBPα, these results led to the hypothesis that Trib1 may act as an E2-conjugating enzyme itself. Secondly, TRIB1 also co-expressed with transmembrane protein TMEM181, which is found to be upregulated after bacterial infection, that brings to hypothesize that TRIB1 recruits E3 ligase to ubiquitinate TMEM181 co-expressed genes like SYLT3 to activate the cell- Chapter 5: Conclusions and Future Work 139 signalling pathways, which further activates adaptive immunity specific to an M2 phenotype. However, this hypothesis requires further validation. On the other hand, co-expression analysis performed on different types of cancer showed

that TRIB1 co-expressed with 28 different immediate-early response (IER) genes in 19 out of 28 cancer datasets interrogated. Amongst all, EGR1 and FOS were the only two genes to co-express with TRIB1 in all 19 cancer datasets, which were also tested through in vitro experiments in different cancer cell lines, in order to validate our initial computational observations. Overall, we found that TRIB1's expression was negatively correlated with that of EGR1 and FOS in both DU145 and PC-3 cell lines (i.e. TRIB1 was higher in PC-3 cells than in DU145, while the opposite was observed for both EGR1 and FOS, whose expression levels were elevated in DU145, compared to PC-3 cells). Further on, EGR1 and FOS were detected to be up-regulated in TRIB1 overexpressed DU145 cells, compared to controls, which suggested TRIB1 may regulate these two IER genes. Moreover, upon stimulating IER genes in HEK293T cells, we observed that TRIB1's expression pattern was similar to that of IER genes which, together with the above results indicate that TRIB1 could also be a member of IER gene family. What is more, TRIB1 co-expressed with differentially expressed IER genes involved in cell signalling and cell cycle related pathways, which highlighted yet another potential role for TRIB1, as a positive regulator of cell polarization and cell migration. Considering many observations were recorded from different types of cell lines, performed at different times and in slightly varying conditions, future tests could include replicates of the same conditions across all cell lines, in order to ensure a more appropriate comparison. For example, both TRIB1 knock-down and overexpression can be induced in the two cancer cell lines DU145 and PC-3, as well as in non-cancerous HEK293 cells and macrophages cell-lines. This would enable us to further investigate the variation in expression levels of TRIB1 between different conditions, which was previously observed computationally. Moreover, by studying the expression levels of all TRIB1, EGR1 and FOS in these samples, we would better understand the interaction and direction of regulation between these 3 genes, as well as whether these are cell-type specific. In addition, RNAseq and coexpression analyses of all these conditions would help us confirm our preliminary results, as well as find other genes of interest that would further highlight the genomic network of TRIB1. All in one, our results suggests that SNPs in 3' TRIB1 are not linked to allele-specific expression. However, they are responsible in altering miRNA binding. More that that, we have shown that expression level of TRIB1 increases with increase in number of SNPs in it's 3' UTR, which suggests that miRNAs are not only responsible in altering the expression of TRIB1, but other factors such as RNA-Binding proteins (RBPs), translational repressors, splicing factors, and riboswitches sites altered due the SNPs could also regulate the expression of TRIB1. Moreover, we also shown that TRIB1 is Chapter 5: Conclusions and Future Work 140 coexpressed with IER genes in 18 out of 28 cancer datasets, which led us to hypothesize that TRIB1 could be a member of IER genes. This might explain the reason why studies on the genome-wide immunity response in TRIB1-knockout mice, carried out by collaborators in the TRAIN consortium failed to find large transcriptional effects linked to TRIB1 as

these were carried out in resting, steadystate cells. Furthermore, it also explains the reason of SNPs in 3’ UTR of TRIB1 is not linked to ASE, as TRIB1 was found low expressed in unstimulated and M1-like macrophages because the RNA data was collected post 2 hours from Salmonella typhimurium infected macrophages (M1-like macrophages). Furthermore, we found SNPs in 3’ UTR altered binding sites of 9 miRNAs, which were expressed in Mun and MLPS+IFNϒ small RNA-seq dataset out of which 6 of these miRNAs had novel binding sites created in the 3’UTR of TRIB1, while 3 had existing sites degraded. As the expression of TRIB1 was found to be low in Mun and MLPS+IFNϒ RNA-seq datasets, that could be why these miRNAs could have no effect on regulating TRIB1. Considering all above observations, the future experiment could be stimulating IER genes in unstimulated, M1-like and M2-like macrophages to record the changes in expression at different time points of TRIB1 and IER genes and change in expression of miRNAs targeting TRIB1 and IER genes in macrophages

References

Agarwal, Vikram, George W. Bell, Jin Wu Nam, and David P. Bartel. 2015. 'Predicting Effective MicroRNA Target Sites in Mammalian MRNAs'. ELife.

Alberts, Bruce, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter. 2017. Molecular Biology of the Cell.

Arndt, Lilli, Janine Dokas, Martin Gericke, Carl Elias Kutzner, Silvana Müller, Franziska Jeromin, Joachim Thiery, and Ralph Burkhardt. 2018. 'Tribbles Homolog 1 Deficiency Modulates Function and Polarization of Murine Bone Marrow– Derived Macrophages'. Journal of Biological Chemistry 293(29):11527–36.

Ashton-Chess, Joanna, Magali Giral, Michael Mengel, Karine Renaudin, Yohann Foucher, Wilfried Gwinner, Christophe Braud, Emilie Dugast, Thibaut Quillard, Pamela Thebault, Elise Chiffoleau, Cécile Braudeau, Béatrice Charreau, Jean Paul Soulillou, and Sophie Brouard. 2008. 'Tribbles-1 as a Novel Biomarker of Chronic Antibody-Mediated Rejection'. Journal of the American Society of Nephrology.

Atkinson, Taylor J., and Marc S. Halfon. 2014. 'Regulation of Gene Expression in the Genomic Context'. Computational and Structural Biotechnology Journal 9(13):e201401001.

Bahn, Jae Hoon, Jae Hyung Lee, Gang Li, Christopher Greer, Guangdun Peng, and Xinshu Xiao. 2012. 'Accurate Identification of A-to-I RNA Editing in Human by Transcriptome Sequencing'. Genome Research.

Bahrami, Shahram, and Finn Drabløs. 2016. 'Gene Regulation in the Immediate-Early Response Process'. Advances in Biological Regulation.

Baidžajevas, Kajus, Éva Hadadi, Bernett Lee, Josephine Lum, Foo Shihui, Ian Sudbery, Endre Kiss-Tóth, Siew Cheng Wong, and Heather L. Wilson. 2020. 'Macrophage Polarisation Associated with Atherosclerosis Differentially Affects Their Capacity to Handle Lipids'. Atherosclerosis.

Bailey, Fiona P., Dominic P. Byrne, Krishnadev Oruganty, Claire E. Eyers, Christopher J. Novotny, Kevan M. Shokat, Natarajan Kannan, and Patrick A. Eyers. 2015. 'The Tribbles 2 (TRB2) Pseudokinase Binds to ATP and Autophosphorylates in a Metal-Independent Manner'. Biochemical Journal 467(1):47–62.

Ballouz, S., W. Verleyen, and J. Gillis. 2015. 'Guidance for RNA-Seq Co-Expression Network Construction and Analysis: Safety in Numbers'. Bioinformatics 31(13):2123–30.

Bauer, Robert C., Makoto Sasaki, Daniel M. Cohen, Jian Cui, Mikhaila A. Smith, Batuhan O. Yenilmez, David J. Steger, and Daniel J. Rader. 2015. 'Tribbles-1 Regulates Hepatic Lipogenesis through Posttranscriptional Regulation of C/EBPα'. Journal of Clinical Investigation 125(10):3809–18.

Belacel, Nabil, Qian Wang, and Miroslava Cuperlovic-Culf. 2006. 'Clustering Methods for Microarray Gene Expression Data'. OMICS A Journal of Integrative Biology 10(4):507–31.

Betel, Doron, Anjali Koppal, Phaedra Agius, Chris Sander, and Christina Leslie. 2010. 'Comprehensive Modeling of MicroRNA Targets Predicts Functional Non-Conserved and Non-Canonical Sites'. Genome Biology.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. Bioinformatics 30(15):2114–20.

Boudeau, Jérôme, Diego Miranda-Saavedra, Geoffrey J. Barton, and Dario R. Alessi. 2006. 'Emerging Roles of Pseudokinases'. Trends in Cell Biology 16(9):443–52.

Brazhnik, Paul, Alberto De La Fuente, and Pedro Mendes. 2002. 'Gene Networks: How to Put the Function in Genomics'. Trends in Biotechnology 20(11):467–72.

Breitling, Rainer, Yang Li, Bruno M. Tesson, Jingyuan Fu, Chunlei Wu, Tim Wiltshire, Alice Gerrits, Leonid V. Bystrykh, Gerald De Haan, Andrew I. Su, and Ritsert C. Jansen. 2008. 'Genetical Genomics: Spotlight on QTL Hotspots'. PLoS Genetics 4(10):2–5.

Browning, Sharon R., and Brian L. Browning. 2011. 'Haplotype Phasing: Existing Methods and New Developments'. Nature Reviews Genetics.

Bryois, Julien, Alfonso Buil, David M. Evans, John P. Kemp, Stephen B. Montgomery, Donald F. Conrad, Karen M. Ho, Susan Ring, Matthew Hurles, Panos Deloukas, George Davey Smith, and Emmanouil T. Dermitzakis. 2014. 'Cis and Trans Effects of Human Genomic Variants on Gene Expression'. PLoS Genetics.

Buckland, Paul R. 2004. 'Allele-Specific Gene Expression Differences in Humans'. Human Molecular Genetics.

Cao, Mu Shui, Bing Ya Liu, Wen Tao Dai, Wei Xin Zhou, Yi Xue Li, and Yuan Yuan Li. 2015. 'Differential Network Analysis Reveals Dysfunctional Regulatory Networks in Gastric Carcinogenesis'. American Journal of Cancer Research.

Castel, Stephane E., Ami Levy-Moonshine, Pejman Mohammadi, Eric Banks, and Tuuli Lappalainen. 2015. 'Tools and Best Practices for Data Processing in Allelic Expression Analysis'. Genome Biology 16(1):1–12.

Caussy, Cyrielle, Sybil Charrière, Christophe Marçais, Mathilde Di Filippo, Agnès Sassolas, Mireille Delay, Vanessa Euthine, Audrey Jalabert, Etienne Lefai, Sophie Rome, and Philippe Moulin. 2014. 'An APOA5 3' UTR Variant Associated with Plasma Triglycerides Triggers Apoa5 Downregulation by Creating a Functional MiR-485-5p Binding Site'. American Journal of Human Genetics 94(1):129–34.

Chan, Mun Chun, Peter H. Nguyen, Brandi N. Davis, Nobumichi Ohoka, Hidetoshi Hayashi, Keyong Du, Giorgio Lagna, and Akiko Hata. 2007. 'A Novel Regulatory Mechanism of the Bone Morphogenetic Protein (BMP) Signaling Pathway Involving the Carboxyl-Terminal Tail Domain of BMP Type II Receptor'. Molecular and Cellular Biology.

Chang, Jessica, Yiqi Zhou, Xiaoli Hu, Lucia Lam, Cameron Henry, Erin M. Green, Ryosuke Kita, Michael S. Kobor, and Hunter B. Fraser. 2013. 'The Molecular Mechanism of a Cis-Regulatory Adaptation in Yeast'. PLoS Genetics.

Chaplin, David D. 2010. 'Overview of the Immune Response'. Journal of Allergy and Clinical Immunology.

Che, J. J., Y. X. Shao, and G. P. Li. 2014. 'Association between Rs1049673 Polymorphism in CD36 and Premature Coronary Heart Disease'. Genetics and Molecular Research 13(3):7708–17.

Chen, Linlin, Huidan Deng, Hengmin Cui, Jing Fang, Zhicai Zuo, Junliang Deng, Yinglun Li, Xun Wang, and Ling Zhao. 2018. 'Inflammatory Responses and Inflammation-Associated Diseases in Organs'. Oncotarget.

Choi, Jinyoung, Sang Taek Kim, and Joe Craft. 2012. 'The Pathogenesis of Systemic Lupus Erythematosus-an Update'. Current Opinion in Immunology.

Christie, Mary, Andreas Boland, Eric Huntzinger, Oliver Weichenrieder, and Elisa Izaurralde. 2013. 'Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylase and the GW182 Proteins'. Molecular Cell 51(3):360–73.

Clop, Alex, Fabienne Marcq, Haruko Takeda, Dimitri Pirottin, Xavier Tordoir, Bernard Bibé, Jacques Bouix, Florian Caiment, Jean Michel Elsen, Francis Eychenne, Catherine Larzul, Elisabeth Laville, Françoise Meish, Dragan Milenkovic, James Tobin, Carole Charlier, and Michel Georges. 2006. 'A Mutation Creating a Potential Illegitimate MicroRNA Target Site in the Myostatin Gene Affects Muscularity in Sheep'. Nature Genetics.

Cooper, Geoffrey M., and Robert E. Hausman. 2007. The Cell: A Molecular Approach 2nd Edition.

Corliss, Bruce A., Mohammad S. Azimi, Jennifer M. Munson, Shayn M. Peirce, and Walter L. Murfee. 2016. 'Macrophages: An Inflammatory Link Between Angiogenesis and Lymphangiogenesis'. Microcirculation.

Cui, Xiangqin, and Gary A. Churchill. 2003. 'Statistical Tests for Differential Expression in CDNA Microarray Experiments'. Genome Biology 4(4).

Curtale, Graziella, Marcello Rubino, and Massimo Locati. 2019. 'MicroRNAs as Molecular Switches in Macrophage Activation'. Frontiers in Immunology.

Das, Vishal, Sourya Bhattacharya, Channakeshavaiah Chikkaputtaiah, Saugata Hazra, and Mintu Pal. 2019. 'The Basics of Epithelial–Mesenchymal Transition (EMT): A Study from a Structure, Dynamics, and Functional Perspective'. Journal of Cellular Physiology.

Depristo, Mark A., Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, Guillermo Del Angel, Manuel A. Rivas, Matt Hanna, Aaron McKenna, Tim J. Fennell, Andrew M. Kernytsky, Andrey Y. Sivachenko, Kristian Cibulskis, Stacey B. Gabriel, David Altshuler, and Mark J. Daly. 2011. 'A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data'. Nature Genetics.

Dimas, Antigone S., Samuel Deutsch, Barbara E. Stranger, Stephen B. Montgomery, Christelle Borel, Homa Attar-Cohen, Catherine Ingle, Claude Beazley, Maria Gutierrez Arcelus, Magdalena Sekowska, Marilyne Gagnebin, James Nisbett, Panos Deloukas, Emmanouil T. Dermitzakis, and Stylianos E. Antonarakis. 2009. 'Common Regulatory Variation Impacts Gene Expression in a Cell Type-Dependent Manner'. Science.

Ding, Jun, Xiaoman Li, and Haiyan Hu. 2016. 'TarPmiR: A New Approach for MicroRNA Target Site Prediction'. Bioinformatics.

Do, Jin Hwan, and Dong Kug Choi. 2008. 'Clustering Approaches to Identifying Gene Expression Patterns from DNA Microarray Data'. Molecules and Cells 25(2):279–88.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. 'STAR: Ultrafast Universal RNA-Seq Aligner'. Bioinformatics 29(1):15–21.

van Dongen, Stijn, Cei Abreu-Goodger, and Anton J. Enright. 2008. 'Detecting MicroRNA Binding and SiRNA Off-Target Effects from Expression Data'. Nature Methods.

Duan, Jubao, Mark S. Wainwright, Josep M. Comeron, Naruya Saitou, Alan R. Sanders, Joel Gelernter, and Pablo V. Gejman. 2003. 'Synonymous Mutations in the Human Dopamine Receptor D2 (DRD2) Affect MRNA Stability and Synthesis of the Receptor'. Human Molecular Genetics.

Dugast, Emilie, Endre Kiss-Toth, Louise Docherty, Richard Danger, Mélanie Chesneau, Virginie Pichard, Jean Paul Judor, Ségolène Pettré, Sophie Conchon, Jean Paul Soulillou, Sophie Brouard, and Joanna Ashton-Chess. 2013. 'Identification of Tribbles-1 as a Novel Binding Partner of Foxp3 in Regulatory T Cells'. Journal of Biological Chemistry.

Eder, Katalin, Hongtao Guan, Hye Y. Sung, Jon Ward, Adrienn Angyal, Michelle Janas, Gabriella Sarmay, Erno Duda, Martin Turner, Steven K. Dower, Sheila E. Francis, David C. Crossman, and Endre Kiss-Toth. 2008. 'Tribbles-2 Is a Novel Regulator of Inflammatory Activation of Monocytes'. International Immunology.

Eisen, Michael B., Paul T. Spellman, Patrick O. Brown, and David Botstein. 1998. 'Cluster Analysis and Display of Genome-Wide Expression Patterns'. Proceedings of the National Academy of Sciences of the United States of America.

Enright, Anton J., Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander, and Debora S. Marks. 2003. 'Enright-2003-Genomebiol.Pdf'. Genome Biology 5(1):R1.

Essandoh, Kobina, Yutian Li, Jiuzhou Huo, and Guo Chang Fan. 2016. 'MiRNA-Mediated Macrophage Polarization and Its Potential Role in the Regulation of Inflammatory Response'. Shock.

Eyers, Patrick A., Karen Keeshan, and Natarajan Kannan. 2017. 'Tribbles in the 21st Century: The Evolving Roles of Tribbles Pseudokinases in Biology and Disease'. Trends in Cell Biology.

Fan, Jiaxin, Jian Hu, Chenyi Xue, Hanrui Zhang, Katalin Susztak, Muredach P. Reilly, Muredach P. Reilly, Rui Xiao, and Mingyao Li. 2020. 'ASEP: Gene-Based Detection of Allele-Specific Expression across Individuals in a Population by RNA Sequencing'. PLoS Genetics 16(5):1–23.

Fang, Zhuo, and Nikolaus Rajewsky. 2011. 'The Impact of MiRNA Target Sites in Coding Sequences and in 3'UTRs'. PLoS ONE.

François Aguet, Kristin G. Ardlie, Tuuli Lappalainen. 2020. The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues.

Galvani, Alison P., and Montgomery Slatkin. 2003. 'Evaluating Plague and Smallpox as Historical Selective Pressures for the CCR5-Δ32 HIV-Resistance Allele'. Proceedings of the National Academy of Sciences of the United States of America 100(25):15276–79.

Garcia, David M., Daehyun Baek, Chanseok Shin, George W. Bell, Andrew Grimson, and David P. Bartel. 2010. 'Weak Seed-Pairing Stability and High Target-Site Abundance Decrease the Proficiency of Lsy-6 and Other MicroRNAs'. Nature Structural and Molecular Biology.

Garrison, Erik, and Gabor Marth. 2012. ‘Haplotype-Based Variant Detection from Short-Read Sequencing -- Free Bayes -- Variant Calling -- Longranger’. ArXiv Preprint ArXiv:1207.3907.

Gartner, Jared J., Stephen C. J. Parker, Todd D. Prickett, Ken Dutton-Regester, Michael L. Stitzel, Jimmy C. Lin, Sean Davis, Vijaya L. Simhadri, Sujata Jha, Nobuko Katagiri, Valer Gotea, Jamie K. Teer, Xiaomu Wei, Mario A. Morken, Umesh K. Bhanot, Guo Chen, Laura L. Elnitski, Michael A. Davies, Jeffrey E. Gershenwald, Hannah Carter, Rachel Karchin, William Robinson, Steven Robinson, Steven A. Rosenberg, Francis S. Collins, Giovanni Parmigiani, Anton A. Komar, Chava Kimchi-Sarfaty, Nicholas K. Hayward, Elliott H. Margulies, and Yardena Samuels. 2013. ‘Whole-Genome Sequencing Identifies a Recurrent Functional Synonymous Mutation in Melanoma’. Proceedings of the National Academy of Sciences of the United States of America.

Geng, Peiliang, Xiaoxin Zhao, Lisha Xiang, Yunmei Liao, Ning Wang, Juanjuan Ou, Ganfeng Xie, Chen Liu, Jianjun Li, Hongtao Li, and Houjie Liang. 2014. ‘Distinct Role of CD86 Polymorphisms (Rs1129055, Rs17281995) In Risk of Cancer: Evidence from a Meta- Analysis’. PLoS ONE.

Gennarino, Vincenzo Alessandro, Giovanni D’Angelo, Gopuraja Dharmalingam, Serena Fernandez, Giorgio Russolillo, Remo Sanges, Margherita Mutarelli, Vincenzo Belcastro, Andrea Ballabio, Pasquale Verde, Marco Sardiello, and Sandro Banfi. 2012. ‘Identification of MicroRNA-Regulated Gene Networks by Expression Analysis of Target Genes’. Genome Research.

Ghanbari, Mohsen, Oscar H. Franco, Hans W. J. De Looper, Albert Hofman, Stefan J. Erkeland, and Abbas Dehghan. 2015. ‘Genetic Variations in MicroRNA-Binding Sites Affect MicroRNA-Mediated Regulation of Several Genes Associated with Cardio-Metabolic Phenotypes’. Circulation: Cardiovascular Genetics.

Ginhoux, Florent, and Steffen Jung. 2014. ‘Monocytes and Macrophages: Developmental Pathways and Tissue Homeostasis’. Nature Reviews Immunology.

Gitenay, Delphine, and Véronique T. Baron. 2009. ‘Is EGR1 a Potential Target for Prostate Cancer Therapy?’ Future Oncology.

Göring, Harald H. H., Joanne E. Curran, Matthew P. Johnson, Thomas D. Dyer, Jac Charlesworth, Shelley A. Cole, Jeremy B. M. Jowett, Lawrence J. Abraham, David L. Rainwater, Anthony G. Comuzzie, Michael C. Mahaney, Laura Almasy, Jean W. MacCluer, Ahmed H. Kissebah, Gregory R. Collier, Eric K. Moses, and John Blangero. 2007. ‘Discovery of Expression QTLs Using Large-Scale Transcriptional Profiling in Human Lymphocytes’. Nature Genetics.

Griesemer, Dustin, James R. Xue, Steve K. Reilly, Jacob C. Ulirsch, Kalki Kukreja, Joe Davis, Masahiro Kanai, David K. Yang, Stephen B. Montgomery, Carl D. Novina, Ryan Tewhey, and Pardis C. Sabeti. 2021. ‘Genome-Wide Functional Screen of 3’UTR Variants Uncovers Causal Variants for Human Disease and Evolution’. SSRN Electronic Journal.

Grimson, Andrew, Kyle Kai-how Farh, Wendy K. Johnston, Philip Garrett-, Lee P. Lim, and David P. Bartel. 2013. ‘Nihms515910’. 27(1):91–105.

Grimson, Andrew, Kyle Kai How Farh, Wendy K. Johnston, Philip Garrett-Engele, Lee P. Lim, and David P. Bartel. 2007. ‘MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing’. Molecular Cell.

de Groot, Amber E., and Kenneth J. Pienta. 2018. 'Epigenetic Control of Macrophage Polarization: Implications for Targeting Tumor-Associated Macrophages'. Oncotarget.

Grundberg, Elin, Kerrin S. Small, Åsa K. Hedman, Alexandra C. Nica, Alfonso Buil, Sarah Keildson, Jordana T. Bell, Tsun Po Yang, Eshwar Meduri, Amy Barrett, James Nisbett, Magdalena Sekowska, Alicja Wilk, So Youn Shin, Daniel Glass, Mary Travers, Josine L. Min, Sue Ring, Karen Ho, Gudmar Thorleifsson, Augustine Kong, Unnur Thorsteindottir, Chrysanthi Ainali, Antigone S. Dimas, Neelam Hassanali, Catherine Ingle, David Knowles, Maria Krestyaninova, Christopher E. Lowe, Paola Di Meglio, Stephen B. Montgomery, Leopold Parts, Simon Potter, Gabriela Surdulescu, Loukia Tsaprouni, Sophia Tsoka, Veronique Bataille, Richard Durbin, Frank O. Nestle, Stephen O'Rahilly, Nicole Soranzo, Cecilia M. Lindgren, Krina T. Zondervan, Kourosh R. Ahmadi, Eric E. Schadt, Kari Stefansson, George Davey Smith, Mark I. McCarthy, Panos Deloukas, Emmanouil T. Dermitzakis, and Tim D. Spector. 2012. 'Mapping Cis-and Trans-Regulatory Effects across Multiple Tissues in Twins'. Nature Genetics 44(10):1084–89.

Guo, Xu, Deyang Li, Yibing Chen, Jiaze An, Kan Wang, Zhuding Xu, Zhinan Chen, and Jinliang Xing. 2015. 'SNP Rs2057482 in HIF1A Gene Predicts Clinical Outcome of Aggressive Hepatocellular Carcinoma Patients after Surgery'. Scientific Reports.

Hackenberg, Michael, Martin Sturm, David Langenberger, Juan Manuel Falcón-Pérez, and Ana M. Aransay. 2009. 'MiRanalyzer: A MicroRNA Detection and Analysis Tool for next-Generation Sequencing Experiments'. Nucleic Acids Research.

Harvey, Chris T., Gregory A. Moyerbrailean, Gordon O. Davis, Xiaoquan Wen, Francesca Luca, and Roger Pique-Regi. 2015. 'QuASAR: Quantitative Allele-Specific Analysis of Reads'. Bioinformatics 31(8):1235–42.

Hegedus, Z., A. Czibula, and E. Kiss-Toth. 2006. 'Tribbles: Novel Regulators of Cell Function; Evolutionary Aspects'. Cellular and Molecular Life Sciences.

Hicks, Steven D., and Frank A. Middleton. 2016. 'A Comparative Review of MicroRNA Expression Patterns in Autism Spectrum Disorder'. Frontiers in Psychiatry.

Hodgkinson, Alan, Jean Christophe Grenier, Elias Gbeha, and Philip Awadalla. 2016. 'A Haplotype-Based Normalization Technique for the Analysis and Detection of Allele Specific Expression'. BMC Bioinformatics.

Hoeffel, Guillaume, Yilin Wang, Melanie Greter, Peter See, Pearline Teo, Benoit Malleret, Marylène Leboeuf, Donovan Low, Guillaume Oller, Francisca Almeida, Sharon H. Y. Choy, Marcos Grisotto, Laurent Renia, Simon J. Conway, E. Richard Stanley, Jerry K. Y. Chan, Lai Guan Ng, Igor M. Samokhvalov, Miriam Merad, and Florent Ginhoux. 2012. 'Adult Langerhans Cells Derive Predominantly from Embryonic Fetal Liver Monocytes with a Minor Contribution of Yolk Sac-Derived Macrophages'. Journal of Experimental Medicine.

Hu, Sen Lin, Guang Lin Cui, Jin Huang, Jian Gang Jiang, and Dao Wen Wang. 2016. 'An APOC3 3'UTR Variant Associated with Plasma Triglycerides Levels and Coronary Heart Disease by Creating a Functional MiR-4271 Binding Site'. Scientific Reports 6(September):1–10.

Hudson, Nicholas J., Brian P. Dalrymple, and Antonio Reverter. 2012. 'Beyond Differential Expression: The Quest for Causal Mutations and Effector Molecules'. BMC Genomics.

Huggins, Christopher J., Manasi K. Mayekar, Nancy Martin, Karen L. Saylor, Mesfin Gonit, Parthav Jailwala, Manjula Kasoji, Diana C. Haines, Octavio A. Quiñones, and Peter F. Johnson. 2016. 'C/EBPγ Is a Critical Regulator of Cellular Stress Response Networks through Heterodimerization with ATF4'. Molecular and Cellular Biology.

Hume, David A., and Kelli P. A. MacDonald. 2012. 'Therapeutic Applications of Macrophage Colony-Stimulating Factor-1 (CSF-1) and Antagonists of CSF-1 Receptor (CSF-1R) Signaling'. Blood.

Humphrey, Rohan K., Christina J. Newcomb, Shu Mei A. Yu, Ergeng Hao, Doris Yu, Stan Krajewski, Keyong Du, and Ulupi S. Jhala. 2010. 'Mixed Lineage Kinase-3 Stabilizes and Functionally Cooperates with TRIBBLES-3 to Compromise Mitochondrial Integrity in Cytokine-Induced Death of Pancreatic Beta Cells'. Journal of Biological Chemistry.

Huntzinger, Eric, and Elisa Izaurralde. 2011. 'Gene Silencing by MicroRNAs: Contributions of Translational Repression and MRNA Decay'. Nature Reviews Genetics 12(2):99–110.

Hye, Youn Sung, Hongtao Guan, Agnes Czibula, Andrea R. King, Katalin Eder, Emily Heath, Kim Suvarna, Steven K. Dower, Anthony G. Wilson, Sheila E. Francis, David C. Crossman, and Endre Kiss-Toth. 2007. 'Human Tribbles-1 Controls Proliferation and Chemotaxis of Smooth Muscle Cells via MAPK Signaling Pathways'. Journal of Biological Chemistry.

Italiani, Paola, and Diana Boraschi. 2014. 'From Monocytes to M1/M2 Macrophages: Phenotypical vs. Functional Differentiation'. Frontiers in Immunology 5(OCT):1–22.

Jacinta-Fernandes, Ana, Joana M. Xavier, Ramiro Magno, Joel G. Lage, and Ana Teresa Maia. 2020. 'Allele-Specific MiRNA-Binding Analysis Identifies Candidate Target Genes for Breast Cancer Risk'. Npj Genomic Medicine.

Jadhav, Kavita S., and Robert C. Bauer. 2019. 'Trouble with Tribbles-1: Elucidating the Mechanism of a Genome-Wide Association Study Locus'. Arteriosclerosis, Thrombosis, and Vascular Biology.

Johnston, Jessica M., Adrienn Angyal, Robert C. Bauer, Stephen Hamby, S. Kim Suvarna, Kajus Baidžajevas, Zoltan Hegedus, T. Neil Dear, Martin Turner, Heather L. Wilson, Alison H. Goodall, Daniel J. Rader, Carol C. Shoulders, Sheila E. Francis, and Endre Kiss-Toth. 2019. 'Myeloid Tribbles 1 Induces Early Atherosclerosis via Enhanced Foam Cell Expansion'. Science Advances 5(10).

Jonas, Stefanie, and Elisa Izaurralde. 2015. 'Towards a Molecular Understanding of MicroRNA-Mediated Gene Silencing'. Nature Reviews Genetics 16(7):421–33.

Jung, Seulgi, Wenting Liu, Jiwon Baek, Jung Won Moon, Byong Duk Ye, Ho Su Lee, Sang Hyoung Park, Suk Kyun Yang, Buhm Han, Jianjun Liu, and Kyuyoung Song. 2020. 'Expression Quantitative Trait Loci (EQTL) Mapping in Korean Patients With Crohn's Disease and Identification of Potential Causal Genes Through Integration With Disease Associations'. Frontiers in Genetics.

Karolchik, Donna, Angela S. Hinricks, Terrence S. Furey, Krishna M. Roskin, Charles W. Sugnet, David Haussler, and W. James Kent. 2004. 'The UCSC Table Browser Data Retrieval Tool'. Nucleic Acids Research.

Kato, Satomi, and Keyong Du. 2007. 'TRB3 Modulates C2C12 Differentiation by Interfering with Akt Activation'. Biochemical and Biophysical Research Communications.

Kaur, Bani Preet, and Elizabeth Secord. 2019. 'Innate Immunity'. Pediatric Clinics of North America.

Keeshan, Karen, Will Bailis, Priya H. Dedhia, Maria E. Vega, Olga Shestova, Lanwei Xu, Kristin Toscano, Sacha N. Uljon, Stephen C. Blacklow, and Warren S. Pear. 2010. 'Transformation by Tribbles Homolog 2 (Trib2) Requires Both the Trib2 Kinase Domain and COP1 Binding'. Blood.

Kim-Hellmuth, Sarah, Matthias Bechheim, Benno Pütz, Pejman Mohammadi, Yohann Nédélec, Nicholas Giangreco, Jessica Becker, Vera Kaiser, Nadine Fricker, Esther Beier, Peter Boor, Stephane E. Castel, Markus M. Nöthen, Luis B. Barreiro, Joseph K. Pickrell, Bertram Müller-Myhsok, Tuuli Lappalainen, Johannes Schumacher, and Veit Hornung. 2017. 'Genetic Regulatory Effects Modified by Immune Activation Contribute to Autoimmune Disease Associations'. Nature Communications.

Kiss-Toth, Endre, Stephanie M. Bagstaff, Hye Y. Sung, Veronika Jozsa, Clare Dempsey, Jim C. Caunt, Kevin M. Oxley, David H. Wyllie, Timea Polgar, Mary Harte, Luke A. J. O'Neil, Eva E. Qwarnstrom, and Steven E. Dower. 2004. 'Human Tribbles, a Protein Family Controlling Mitogen-Activated Protein Kinase Cascades'. Journal of Biological Chemistry.

Kogelman, Lisette J. A., and Haja N. Kadarmideen. 2014. 'Weighted Interaction SNP Hub (WISH) Network Method for Building Genetic Networks for Complex Diseases and Traits Using Whole Genome Genotype Data'. BMC Systems Biology 8(Suppl 2):S5.

Kohonen, Teuvo. 1982. 'Self-Organized Formation of Topologically Correct Feature Maps'. Biological Cybernetics.

Komada, Takanori, and Daniel A. Muruve. 2019. 'The Role of Inflammasomes in Kidney Disease'. Nature Reviews Nephrology.

Koo, Seung Hoi, Hiroaki Satoh, Stephan Herzig, Chih Hao Lee, Susan Hedrick, Rohit Kulkarni, Ronald M. Evans, Jerrold Olefsky, and Marc Montminy. 2004. 'PGC-1 Promotes Insulin Resistance in Liver through PPAR-α-Dependent Induction of TRB-3'. Nature Medicine.

Kozomara, Ana, Maria Birgaoanu, and Sam Griffiths-Jones. 2019. 'MiRBase: From MicroRNA Sequences to Function'. Nucleic Acids Research.

Krek, Azra, Dominic Grün, Matthew N. Poy, Rachel Wolf, Lauren Rosenberg, Eric J. Epstein, Philip MacMenamin, Isabelle Da Piedade, Kristin C. Gunsalus, Markus Stoffel, and Nikolaus Rajewsky. 2005. 'Combinatorial MicroRNA Target Predictions'. Nature Genetics.

Kruse, Jan Philipp, and Wei Gu. 2009. 'Modes of P53 Regulation'. Cell.

Ku, Chee Seng, En Yun Loy, Agus Salim, Yudi Pawitan, and Kee Seng Chia. 2010. 'The Discovery of Human Genetic Variations and Their Use as Disease Markers: Past, Present and Future'. Journal of Human Genetics.

Kwasniak, Konrad, Justyna Czarnik-Kwasniak, Aleksandra Maziarz, David Aebisher, Kinga Zielinska, Bozenna Karczmarek-Borowska, and Jacek Tabarkiewicz. 2019. 'Scientific Reports Concerning the Impact of Interleukin 4, Interleukin 10 and Transforming Growth Factor β on Cancer Cells'. Central European Journal of Immunology.

Langfelder, Peter, and Steve Horvath. 2008. 'WGCNA: An R Package for Weighted Correlation Network Analysis'. BMC Bioinformatics 9.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. 'Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome'. Genome Biology.

Lathrop, Stephanie K., Kendal G. Cooper, Kelsey A. Binder, Tregei Starr, Veena Mampilli, Corrella S. Detweiler, and Olivia Steele-Mortimer. 2018. 'Salmonella Typhimurium Infection of Human Monocyte-Derived Macrophages'. Current Protocols in Microbiology.

Lee, Homin K., Amy K. Hsu, Jon Sajdak, Jie Qin, and Paul Pavlidis. 2004. 'Coexpresion Analysis of Human Genes across Many Microarray Data Sets'. Genome Research 14(6):1085–94.

Lewis, Benjamin P., Christopher B. Burge, and David P. Bartel. 2005. 'Conserved Seed Pairing, Often Flanked by Adenosines, Indicates That Thousands of Human Genes Are MicroRNA Targets'. Cell.

Lewis, Benjamin P., I. Hung Shih, Matthew W. Jones-Rhoades, David P. Bartel, and Christopher B. Burge. 2003. 'Prediction of Mammalian MicroRNA Targets'. Cell.

Li, Heng, and Richard Durbin. 2009. 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform'. Bioinformatics.

Li, Xian Chang, and Laurence A. Turka. 2010. 'An Update on Regulatory T Cells in Transplant Tolerance and Rejection'. Nature Reviews Nephrology.

Liang, Jian Wei, Zhi Zhou Shi, Tong Tong Zhang, Jia Jie Hao, Zheng Wang, Xiao Min Wang, Hai Yang, Ming Rong Wang, Zhi Xiang Zhou, and Yu Zhang. 2013. 'Analysis of Genomic Aberrations Associated with the Clinicopathological Parameters of Rectal Cancer by Array-Based Comparative Genomic Hybridization'. Oncology Reports 29(5):1835–40.

Liang, Weiwei, Fangfang Sun, Yiming Zhao, Lizhen Shan, and Hanyu Lou. 2020. 'Identification of Susceptibility Modules and Genes for Cardiovascular Disease in Diabetic Patients Using WGCNA Analysis'. Journal of Diabetes Research.

Liao, Yi, Yulei Wang, Mengqing Cheng, Chengliang Huang, and Xianming Fan. 2020. 'Weighted Gene Coexpression Network Analysis of Features That Control Cancer Stem Cells Reveals Prognostic Biomarkers in Lung Adenocarcinoma'. Frontiers in Genetics 11(April):1–14.

Lin, Zhuo Yuan, Ya Qiang Huang, Yan Qiong Zhang, Zhao Dong Han, Hui Chan He, Xiao Hui Ling, Xin Fu, Qi Shan Dai, Chao Cai, Jia Hong Chen, Yu Xiang Liang, Fu Neng Jiang, Wei De Zhong, Fen Wang, and Chin Lee Wu. 2014. 'MicroRNA-224 Inhibits Progression of Human Prostate Cancer by Downregulating TRIB1'. International Journal of Cancer.

Liu, Gang, and Edward Abraham. 2013. 'MicroRNAs in Immune Response and Macrophage Polarization'. Arteriosclerosis, Thrombosis, and Vascular Biology.

Liu, Yi Hsia, Karen A. L. Tan, Ivan W. Morrison, Jonathan R. Lamb, and David J. Argyle. 2013. 'Macrophage Migration Is Controlled by Tribbles 1 through the Interaction between C/EBPβ and TNF-α'. Veterinary Immunology and Immunopathology 155(1–2):67–75.

Liu, Zhi, Xiao Dong, and Yixue Li. 2018a. 'A Genome-Wide Study of Allele-Specific Expression in Colorectal Cancer'. Frontiers in Genetics.

Liu, Zhi, Xiao Dong, and Yixue Li. 2018b. 'A Genome-Wide Study of Allele-Specific Expression in Colorectal Cancer'. Frontiers in Genetics.

Lopalco, Lucia. 2010. 'CCR5: From Natural Resistance to a New Anti-HIV Strategy'. Viruses 2(2):574–600.

Lorenz, Ronny, Stephan H. Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. 2011. 'ViennaRNA Package 2.0'. Algorithms for Molecular Biology 6(1):1–14.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. 'Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2'. Genome Biology.

Lu, Rong, Ryan M. Smith, Michal Seweryn, Danxin Wang, Katherine Hartmann, Amy Webb, Wolfgang Sadee, and Grzegorz A. Rempala. 2015. 'Analyzing Allele Specific RNA Expression Using Mixture Models'. BMC Genomics.

Machiela, Mitchell J., and Stephen J. Chanock. 2015. 'LDlink: A Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants'. Bioinformatics.

MacQueen, J. 1967. 'Some Methods for Classification and Analysis of Multivariate Observations'. in Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability.

Maia, Ana Teresa, Inmaculada Spiteri, Alvin J. X. Lee, Martin O'Reilly, Linda Jones, Carlos Caldas, and Bruce A. J. Ponder. 2009. 'Extent of Differential Allelic Expression of Candidate Breast Cancer Genes Is Similar in Blood and Breast'. Breast Cancer Research.

Mantovani, Alberto, Subhra K. Biswas, Maria Rosaria Galdiero, Antonio Sica, and Massimo Locati. 2013. 'Macrophage Plasticity and Polarization in Tissue Repair and Remodelling'. Journal of Pathology.

Martin, Hilary C., Shivangi Wani, Anita L. Steptoe, Keerthana Krishnan, Katia Nones, Ehsan Nourbakhsh, Alexander Vlassov, Sean M. Grimmond, and Nicole Cloonan. 2014. 'Imperfect Centered MiRNA Binding Sites Are Common and Can Mediate Repression of Target MRNAs'. Genome Biology 15(3).

Mason, Mike J., Guoping Fan, Kathrin Plath, Qing Zhou, and Steve Horvath. 2009. 'Signed Weighted Gene Co-Expression Network Analysis of Transcriptional Regulation in Murine Embryonic Stem Cells'. BMC Genomics 10.

Matsumoto, Michihiro, Seongah Han, Tadahiro Kitamura, and Domenico Accili. 2006. 'Dual Role of Transcription Factor FoxO1 in Controlling Hepatic Insulin Sensitivity and Lipid Metabolism'. Journal of Clinical Investigation.

Matzinger, Polly. 2007. 'Friendly and Dangerous Signals: Is the Tissue in Control?' Nature Immunology 8(1):11–13.

Mayba, Oleg, Houston N. Gilbert, Jinfeng Liu, Peter M. Haverty, Suchit Jhunjhunwala, Zhaoshi Jiang, Colin Watanabe, and Zemin Zhang. 2014. 'MBASED: Allele-Specific Expression Detection in Cancer Tissues and Cell Lines'. Genome Biology 15(8):1–21.

Mège, Jean Louis, Vikram Mehraj, and Christian Capo. 2011. 'Macrophage Polarization and Bacterial Infections'. Current Opinion in Infectious Diseases.

Meyer, Patrick E., Frédéric Lafitte, and Gianluca Bontempi. 2008. 'Minet: A r/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information'. BMC Bioinformatics 9:1–10.

Min, Hyeyoung, and Sungroh Yoon. 2010. 'Got Target?: Computational Methods for MicroRNA Target Prediction and Their Extension'. Experimental and Molecular Medicine.

Miyajima, Chiharu, Yasumichi Inoue, and Hidetoshi Hayashi. 2015. 'Pseudokinase Tribbles 1 (TRB1) Negatively Regulates Tumor-Suppressor Activity of P53 through P53 Deacetylation'. Biological and Pharmaceutical Bulletin.

Mohan, Chandra, and Chaim Putterman. 2015. 'Genetics and Pathogenesis of Systemic Lupus Erythematosus and Lupus Nephritis'. Nature Reviews Nephrology.

Mortazavi, Ali, Brian A. Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. 2008. 'Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq'. Nature Methods 5(7):621–28.

Murray, Peter J. 2017. 'Macrophage Polarization'. Annual Review of Physiology 79:541–66.

Naiki, Takahiro, Eiko Saijou, Yuichiro Miyaoka, Keisuke Sekine, and Atsushi Miyajima. 2007. 'TRB2, a Mouse Tribbles Ortholog, Suppresses Adipocyte Differentiation by Inhibiting AKT and C/EBP'. Journal of Biological Chemistry.

Narasimhan, Vagheesh, Petr Danecek, Aylwyn Scally, Yali Xue, Chris Tyler-Smith, and Richard Durbin. 2016. 'BCFtools/RoH: A Hidden Markov Model Approach for Detecting Autozygosity from next-Generation Sequencing Data'. Bioinformatics.

Nau, Gerard J., Joan F. L. Richmond, Ann Schlesinger, Ezra G. Jennings, Eric S. Lander, and Richard A. Young. 2002. 'Human Macrophage Activation Programs Induced by Bacterial Pathogens'. Proceedings of the National Academy of Sciences of the United States of America.

Nédélec, Yohann, Joaquín Sanz, Golshid Baharian, Zachary A. Szpiech, Alain Pacis, Anne Dumaine, Jean Christophe Grenier, Andrew Freiman, Aaron J. Sams, Steven Hebert, Ariane Pagé Sabourin, Francesca Luca, Ran Blekhman, Ryan D. Hernandez, Roger Pique-Regi, Jenny Tung, Vania Yotova, and Luis B. Barreiro. 2016. 'Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens'. Cell 167(3):657--669.e21.

Neph, Shane, M. Scott Kuehn, Alex P. Reynolds, Eric Haugen, Robert E. Thurman, Audra K. Johnson, Eric Rynes, Matthew T. Maurano, Jeff Vierstra, Sean Thomas, Richard Sandstrom, Richard Humbert, and John A. Stamatoyannopoulos. 2012. 'BEDOPS: High-Performance Genomic Feature Operations'. Bioinformatics.

Newbury, S. F. 2006. 'Control of MRNA Stability in Eukaryotes'. in Biochemical Society Transactions.

Ni, Jun, Dan Wang, and Sheng Wang. 2018. 'The CCR5-Delta32 Genetic Polymorphism and HIV-1 Infection Susceptibility: A Meta-Analysis'. Open Medicine (Poland) 13(1):467–74.

Niespolo, Chiara, Jessica M. Johnston, Sumeet R. Deshmukh, Swapna Satam, Ziyanda Shologu, Oscar Villacanas, Ian M. Sudbery, Heather L. Wilson, and Endre Kiss-Toth. 2020. 'Tribbles-1 Expression and Its Function to Control Inflammatory Cytokines, Including Interleukin-8 Levels Are Regulated by MiRNAs in Macrophages and Prostate Cancer Cells'. Frontiers in Immunology 11(November):1–18.

O'Brien, Jacob, Heyam Hayder, Yara Zayed, and Chun Peng. 2018. 'Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation'. Frontiers in Endocrinology.

Okamoto, Haruka, Esther Latres, Rong Liu, Karen Thabet, Andrew Murphy, David Valenzeula, George D. Yancopoulos, Trevor N. Stitt, David J. Glass, and Mark W. Sleeman. 2007. 'Genetic Deletion of Trb3, the Mammalian Drosophila Tribbles Homolog, Displays Normal Hepatic Insulin Signaling and Glucose Homeostasis'. Diabetes.

Okuzaki, Daisuke, Kaori Ota, Shin Ichi Takatsuki, Yukari Akiyoshi, Kazuyuki Naoi, Norikazu Yabuta, Tsutomu Saji, and Hiroshi Nojima. 2017. 'FCN1 (M-Ficolin), Which Directly Associates with Immunoglobulin G1, Is a Molecular Target of Intravenous Immunoglobulin Therapy for Kawasaki Disease'. Scientific Reports.

Ord, Tiit, and Tonis Ord. 2017. 'Mammalian Pseudokinase TRIB3 in Normal Physiology and Disease: Charting the Progress in Old and New Avenues'. Current Protein & Peptide Science.

Ostertag, Anke, Allan Jones, Adam J. Rose, Maria Liebert, Stefan Kleinsorg, Anja Reimann, Alexandros Vegiopoulos, Mauricio Berriel Diaz, Daniela Strzoda, Masahiro Yamamoto, Takashi Satoh, Shizuo Akira, and Stephan Herzig. 2010. 'Control of Adipose Tissue Inflammation through TRB1'. Diabetes.

Pan, Cuiping, Chanchal Kumar, Sebastian Bohl, Ursula Klingmueller, and Matthias Mann. 2009. 'Comparative Proteomic Phenotyping of Cell Lines and Primary Cells to Assess Preservation of Cell Type-Specific Functions'. Molecular and Cellular Proteomics.

Pandey, Ram Vinay, Susanne U. Franssen, Andreas Futschik, and Christian Schlötterer. 2013. 'Allelic Imbalance Metre (Allim), a New Tool for Measuring Allele-Specific Gene Expression with RNA-Seq Data'. Molecular Ecology Resources.

Park, Eunmi, Hyungjin Kim, Jung Min Kim, Benjamin Primack, Sofia Vidal-Cardenas, Ye Xu, Brendan D. Price, Alea A. Mills, and Alan D. D'Andrea. 2013. 'FANCD2 Activates Transcription of TAp63 and Suppresses Tumorigenesis'. Molecular Cell.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, Carl Kingsford, and Computational Biology. 2017. 'Salmon: Fast and Bias-Aware Quantification of Transcript Expression Using Dual-Phase Inference'. Nature Methods.

Pérez, L. Márquez, S. Alonso López, J. L. Herrer. Fajes, and L. Carrión Martín. 2020. 'Hepatocellular Carcinoma'. Medicine (Spain).

Pirim, Harun, Burak Ekşioglu, Andy D. Perkins, and Çetin Yüceer. 2012. 'Clustering of High Throughput Gene Expression Data'. Computers and Operations Research 39(12):3046–61.

Pontes, Beatriz, Raúl Giráldez, and Jesús S. Aguilar-Ruiz. 2015. 'Biclustering on Expression Data: A Review'. Journal of Biomedical Informatics.

Poplin, Ryan, Valentin Ruano-Rubio, Mark A. DePristo, Tim J. Fennell, Mauricio O. Carneiro, Geraldine A. Van der Auwera, David E. Kling, Laura D. Gauthier, Ami Levy-Moonshine, David Roazen, Khalid Shakir, Joel Thibault, Sheila Chandran, Chris Whelan, Monkol Lek, Stacey Gabriel, Mark J. Daly, Ben Neale, Daniel G. MacArthur, and Eric Banks. 2017. 'Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples'. BioRxiv.

Puiffe, Marie Line, Cécile Le Page, Abdelali Filali-Mouhim, Magdalena Zietarska, Véronique Ouellet, Patricia N. Tonin, Mario Chevrette, Diane M. Provencher, and Anne Marie Mes-Masson. 2007. 'Characterization of Ovarian Cancer Ascites on Cell Invasion, Proliferation, Spheroid Formation, and Gene Expression in an in Vitro Model of Epithelial Ovarian Cancer'. Neoplasia.

Puskas, L. G., F. Juhasz, A. Zarva, L. Hackler, and N. R. Farid. 2005. 'Gene Profiling Identifies Genes Specific for Well-Differentiated Epithelial Thyroid Tumors'. in Cellular and Molecular Biology.

Qi, Ling, Jose E. Heredia, Judith Y. Altarejos, Robert Screaton, Naomi Goebel, Sherry Niessen, Ian X. MacLeod, Chong Wee Liew, Rohit N. Kulkarni, James Bain, Christopher Newgard, Michael Welson, Ronald M. Evans, John Yates, and Marc Montminy. 2006. 'TRB3 Links the E3 Ubiquitin Ligase COF1 to Lipid Metabolism'. Science.

Qin, Guangrong, Zhenhao Liu, and Lu Xie. 2021. 'Multiple Omics Data Integration'. in Systems Medicine.

Quinlan, Aaron R., and Ira M. Hall. 2010. 'BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features'. Bioinformatics.

Raj, Towfique, Katie Rothamel, Sara Mostafavi, Chun Ye, Mark N. Lee, Joseph M. Replogle, Ting Feng, Michelle Lee, Natasha Asinovski, Irene Frohlich, Selina Imboywa, Alina Von Korff, Yukinori Okada, Nikolaos A. Patsopoulos, Scott Davis, Cristin McCabe, Hyun Il Paik, Gyan P. Srivastava, Soumya Raychaudhuri, David A. Hafler, Daphne Koller, Aviv Regev, Nir Hacohen, Diane Mathis, Christophe Benoist, Barbara E. Stranger, and Philip L. De Jager. 2014. 'Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes'. Science 344(6183):519–23.

Ramaswami, Gokul, Wei Lin, Robert Piskol, Meng How Tan, Carrie Davis, and Jin Billy Li. 2012. 'Accurate Identification of Human Alu and Non-Alu RNA Editing Sites'. Nature Methods.

Rao, Mohan S., Terry R. Van Vleet, Rita Ciurlionis, Wayne R. Buck, Scott W. Mittelstadt, Eric A. G. Blomme, and Michael J. Liguori. 2019. 'Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver from Short-Term Rat Toxicity Studies'. Frontiers in Genetics.

Rhee, Seung Yon, and Marek Mutwil. 2014. 'Towards Revealing the Functions of All Genes in Plants'. Trends in Plant Science 19(4):212–21.

Ribeiro, António J. M., Sayoni Das, Natalie Dawson, Rossana Zaru, Sandra Orchard, Janet M. Thornton, Christine Orengo, Elton Zeqiraj, James M. Murphy, and Patrick A. Eyers. 2019. 'Emerging Concepts in Pseudoenzyme Classification, Evolution, and Signaling'. Science Signaling 12(594).

Riffo-Campos, Ángela L., Ismael Riquelme, and Priscilla Brebi-Mieville. 2016. 'Tools for Sequence-Based MiRNA Target Prediction: What to Choose?' International Journal of Molecular Sciences 17(12).

Rossi, Marianna N., Antonia Pascarella, Valerio Licursi, Ivan Caiello, Anna Taranta, Laura R. Rega, Elena Levtchenko, Francesco Emma, Fabrizio De Benedetti, and Giusi Prencipe. 2019. 'NLRP2 Regulates Proinflammatory and Antiapoptotic Responses in Proximal Tubular Epithelial Cells'. Frontiers in Cell and Developmental Biology.

Rowczenio, Dorota, Ahmet Dogan, Jason D. Theis, Julie A. Vrana, Helen J. Lachmann, Ashutosh D. Wechalekar, Janet A. Gilbertson, Toby Hunt, Simon D. J. Gibbs, Prayman T. Sattianayagam, Jenny H. Pinney, Philip N. Hawkins, and Julian D. Gillmore. 2011. 'Amyloidogenicity and Clinical Phenotype Associated with Five Novel Mutations in Apolipoprotein A-I'. American Journal of Pathology.

Rozowsky, Joel, Alexej Abyzov, Jing Wang, Pedro Alves, Debasish Raha, Arif Harmanci, Jing Leng, Robert Bjornson, Yong Kong, Naoki Kitabayashi, Nitin Bhardwaj, Mark Rubin, Michael Snyder, and Mark Gerstein. 2011. 'AlleleSeq: Analysis of Allele-Specific Expression and Binding in a Network Framework'. Molecular Systems Biology 7(522):1–15.

Ryckman, Kelli, and Scott M. Williams. 2008. 'Calculation and Use of the Hardy-Weinberg Model in Association Studies'. Current Protocols in Human Genetics.

Salavati, Mazdak, Stephen J. Bush, Sergio Palma-Vera, Mary E. B. McCulloch, David A. Hume, and Emily L. Clark. 2019. 'Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep'. Frontiers in Genetics.

Sathyanarayana, Pradeep, Arvind Dev, Jing Fang, Estelle Houde, Olga Bogacheva, Oleg Bogachev, Madhu Menon, Sarah Browne, Anamika Pradeep, Christine Emerson, and Don M. Wojchowski. 2008. 'EPO Receptor Circuits for Primary Erythroblast Survival'. Blood.

Satoh, Takashi, Hiroyasu Kidoya, Hisamichi Naito, Masahiro Yamamoto, Naoki Takemura, Katsuhiro Nakagawa, Yoshichika Yoshioka, Eiichi Morii, Nobuyuki Takakura, Osamu Takeuchi, and Shizuo Akira. 2013. 'Critical Role of Trib1 in Differentiation of Tissue-Resident M2-like Macrophages'. Nature 495(7442):524–28.

Schwanhüusser, Björn, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. 2011. 'Global Quantification of Mammalian Gene Expression Control'. Nature.

Serin, Elise A. R., Harm Nijveen, Henk W. M. Hilhorst, and Wilco Ligterink. 2016. 'Learning from Co-Expression Networks: Possibilities and Challenges'. Frontiers in Plant Science 7(APR2016):1–18.

Shahrouzi, Parastoo, Ianire Astobiza, Ana R. Cortazar, Verónica Torrano, Alice Macchia, Juana M. Flores, Chiara Niespolo, Isabel Mendizabal, Ruben Caloto, Amaia Ercilla, Laura Camacho, Leire Arreal, Maider Bizkarguenaga, Maria L. Martinez-Chantar, Xose R. Bustelo, Edurne Berra, Endre Kiss-Toth, Guillermo Velasco, Amaia Zabala-Letona, and Arkaitz Carracedo. 2020. 'Genomic and Functional Regulation of TRIB1 Contributes to Prostate Cancer Pathogenesis'. Cancers.

Shankar, Eswar, Kyung Song, Sarah L. Corum, Kara L. Bane, Hui Wang, Hung Ying Kao, and David Danielpour. 2016. 'A Signaling Network Controlling Androgenic Repression of C-Fos Protein in Prostate Adenocarcinoma Cells'. Journal of Biological Chemistry.

Sharova, Lioudmila V., Alexei A. Sharov, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru S. H. Ko. 2009. 'Database for MRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells'. DNA Research.

Soneson, Charlotte, Michael I. Love, and Mark D. Robinson. 2015. 'Differential Analyses for RNA-Seq: Transcript-Level Estimates Improve Gene-Level Inferences'. F1000Research.

Song, Lin, Peter Langfelder, and Steve Horvath. 2012. 'Comparison of Co-Expression Measures: Mutual Information, Correlation, and Model Based Indices'. BMC Bioinformatics 13(1).

Soubeyrand, Sébastien, Amy Martinuk, Paulina Lau, and Ruth McPherson. 2016. 'TRIB1 Is Regulated Post-Transcriptionally by Proteasomal and Non-Proteasomal Pathways'. PLoS ONE.

Soubeyrand, Sébastien, Amy Martinuk, Thet Naing, Paulina Lau, and Ruth McPherson. 2016. 'Role of Tribbles Pseudokinase 1 (TRIB1) in Human Hepatocyte Metabolism'. Biochimica et Biophysica Acta - Molecular Basis of Disease.

Soubeyrand, Sébastien, Thet Naing, Amy Martinuk, and Ruth McPherson. 2013. 'ERK1/2 Regulates Hepatocyte Trib1 in Response to Mitochondrial Dysfunction'. Biochimica et Biophysica Acta - Molecular Cell Research 1833(12):3405–14.

Spellman, Paul T., Gavin Sherlock, Michael Q. Zhang, Vishwanath R. Iyer, Kirk Anders, Michael B. Eisen, Patrick O. Brown, David Botstein, and Bruce Futcher. 1998. 'Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization'. Molecular Biology of the Cell.

Storlazzi, Clelia Tiziana, Thoas Fioretos, Cecilia Surace, Angelo Lonoce, Angela Mastrorilli, Bodil Strömbeck, Pietro D'Addabbo, Francesco Iacovelli, Crescenzio Minervini, Anna Aventin, Nicole Dastugue, Christa Fonatsch, Anne Hagemeijer, Martine Jotterand, Dominique Mühlematter, Marina Lafage-Pochitaloff, Florence Nguyen-Khac, Claudia Schoch, Marilyn L. Slovak, Arabella Smith, Francesc Solè, Nadine Van Roy, Bertil Johansson, and Mariano Rocchi. 2006. 'MYC-Containing Double Minutes in Hematologic Malignancies: Evidence in Favor of the Episome Model and Exclusion of MYC as the Target Gene'. Human Molecular Genetics 15(6):933–42.

Sweeney, Blake A., Anton I. Petrov, Boris Burkov, Robert D. Finn, Alex Bateman, Maciej Szymanski, Wojciech M. Karlowski, Jan Gorodkin, Stefan E. Seemann, Jamie J. Cannone, Robin R. Gutell, Petra Fey, Siddhartha Basu, Simon Kay, Guy Cochrane, Kostantinos Billis, David Emmert, Steven J. Marygold, Rachael P. Huntley, Ruth C. Lovering, Adam Frankish, Patricia P. Chan, Todd M. Lowe, Elspeth Bruford, Ruth Seal, Jo Vandesompele, Pieter Jan Volders, Maria Paraskevopoulou, Lina Ma, Zhang Zhang, Sam Griffiths-Jones, Janusz M. Bujnicki, Pietro Boccaletto, Judith A. Blake, Carol J. Bult, Runsheng Chen, Yi Zhao, Valerie Wood, Kim Rutherford, Elena Rivas, James Cole, Stanley J. F. Laulederkind, Mary Shimoyama, Marc E. Gillespie, Marija Orlic-Milacic, Ioanna Kalvari, Eric Nawrocki, Stacia R. Engel, J. Michael Cherry, Silva Team, Tanya Z. Berardini, Artemis Hatzigeorgiou, Dimitra Karagkouni, Kevin Howe, Paul Davis, Marcel Dinger, Shunmin He, Maki Yoshihama, Naoya Kenmochi, Peter F. Stadler, and Kelly P. Williams. 2019. 'RNAcentral: A Hub of Information for Non-Coding RNA Sequences'. Nucleic Acids Research.

Tauber, Alfred I. 2017. 'Immunity: How Elie Metchnikoff Changed the Course of Modern Medicine by Luba Vikhanski'. Bulletin of the History of Medicine.

Taylor, Susan S., and Alexandr P. Kornev. 2010. 'Yet Another "Active" Pseudokinase, Erb3'. Proceedings of the National Academy of Sciences of the United States of America 107(18):8047–48.

Thiriot, Joseph D., Yazmin B. Martinez-Martinez, Janice J. Endsley, and Alfredo G. Torres. 2020. 'Hacking the Host: Exploitation of Macrophage Polarization by Intracellular Bacterial Pathogens'. Pathogens and Disease.

Tian, Bin, and James L. Manley. 2016. 'Alternative Polyadenylation of MRNA Precursors'. Nature Reviews Molecular Cell Biology.

Turner, Mark D., Belinda Nedjai, Tara Hurst, and Daniel J. Pennington. 2014. 'Cytokines and Chemokines: At the Crossroads of Cell Signalling and Inflammatory Disease'. Biochimica et Biophysica Acta - Molecular Cell Research.

Vinogradov, Alexander E., and Olga V. Anatskaya. 2007. 'Organismal Complexity, Cell Differentiation and Gene Expression: Human over Mouse'. Nucleic Acids Research.

Viola, Antonella, Fabio Munari, Ricardo Sánchez-Rodríguez, Tommaso Scolaro, and Alessandra Castegna. 2019. 'The Metabolic Signature of Macrophage Responses'. Frontiers in Immunology.

Võsa, Urmo, Annique Claringbould, Harm Jan Westra, Marc Jan Bonder, Patrick Deelen, Biao Zeng, Holger Kirsten, Ashis Saha, Roman Kreuzhuber, Silva Kasela, Natalia Pervjakova, Isabel Alvaes, Marie Julie Fave, Mawusse Agbessi, Mark Christiansen, Rick Jansen, Ilkka Seppälä, Lin Tong, Alexander Teumer, Katharina Schramm, Gibran Hemani, Joost Verlouw, Hanieh Yaghootkar, Reyhan Sönmez, Andrew Brown, Viktorija Kukushkina, Anette Kalnapenkis, Sina Rüeger, Eleonora Porcu, Jaanika Kronberg-Guzman, Johannes Kettunen, Joseph Powell, Bernett Lee, Futao Zhang, Wibowo Arindrarto, Frank Beutner, Harm Brugge, Julia Dmitreva, Mahmoud Elansary, Benjamin P. Fairfax, Michel Georges, Bastiaan T. Heijmans, Mika Kähönen, Yungil Kim, Julian C. Knight, Peter Kovacs, Knut Krohn, Shuang Li, Markus Loeffler, Urko M. Marigorta, Hailang Mei, Yukihide Momozawa, Martina Müller-Nurasyid, Matthias Nauck, Michel Nivard, Brenda Penninx, Jonathan Pritchard, Olli Raitakari, Olaf Rotzchke, Eline P. Slagboom, Coen D. A. Stehouwer, Michael Stumvoll, Patrick Sullivan, Peter A. C. 't Hoen, Joachim Thiery, Anke Tönjes, Jenny van Dongen, Maarten van Iterson, Jan Veldink, Uwe Völker, Cisca Wijmenga, Morris Swertz, Anand Andiappan, Grant W. Montgomery, Samuli Ripatti, Markus Perola, Zoltan Kutalik, Emmanouil Dermitzakis, Sven Bergmann, Timothy Frayling, Joyce van Meurs, Holger Prokisch, Habibul Ahsan, Brandon Pierce, Terho Lehtimäki, Dorret Boomsma, Bruce M. Psaty, Sina A. Gharib, Philip Awadalla, Lili Milani, Willem Ouwehand, Kate Downes, Oliver Stegle, Alexis Battle, Jian Yang, Peter M. Visscher, Markus Scholz, Gregory Gibson, Tõnu Esko, and Lude Franke. 2018. 'Unraveling the Polygenic Architecture of Complex Traits Using Blood EQTL Meta-Analysis'. BioRxiv.

Wade, Mark, Yunyuan V. Wang, and Geoffrey M. Wahl. 2010. 'The P53 Orchestra: Mdm2 and Mdmx Set the Tone'. Trends in Cell Biology.

Wang, Anyou, Mufadhal Al-Kuhlani, S. Claiborne Johnston, David M. Ojcius, Joyce Chou, and Deborah Dean. 2013. 'Transcription Factor Complex AP-1 Mediates Inflammation Initiated by Chlamydia Pneumoniae Infection'. Cellular Microbiology 15(5):779–94.

Wang, Heming, Sinuo Chen, Jiayi Wei, Guangqi Song, and Yicheng Zhao. 2021. 'A-to-I RNA Editing in Cancer: From Evaluating the Editing Level to Exploring the Editing Effects'. Frontiers in Oncology.

Wang, Meilin, Mulong Du, Lan Ma, Haiyan Chu, Qiang Lv, Dingwei Ye, Jianming Guo, Chengyuan Gu, Guowei Xia, Yao Zhu, Qiang Ding, Lin Yuan, Guangbo Fu, Na Tong, Chao Qin, Changjun Yin, Jianfeng Xu, and Zhengdong Zhang. 2016. 'A Functional Variant in TP63 at 3q28 Associated with Bladder Cancer Risk by Creating an MiR-140-5p Binding Site'. International Journal of Cancer.

Wang, Yuhui, Nan Wu, Bo Pang, Dandan Tong, Donglin Sun, and Haiming Sun. 2017. 'TRIB1 Promotes Colorectal Cancer Cell Migration and Invasion through Activation MMP-2 via FAK / Src and ERK Pathways'. 8(29):47931–42.

Wen, Xiling, Stefanie Fuhrman, George S. Michaels, Daniel B. Carr, Susan Smith, Jeffery L. Barker, and Roland Somogyi. 1998. 'Large-Scale Temporal Gene Expression Mapping of Central Nervous System Development'. Proceedings of the National Academy of Sciences of the United States of America 95(1):334–39.

Westfall, Matthew D., and Jennifer A. Pietenpol. 2004. 'P63: Molecular Complexity in Development and Cancer'. Carcinogenesis 25(6):857–64.

Wilkin, Françoise, Nathalie Suarez-Huerta, Bernard Robaye, Julien Peetermans, Frédérick Libert, Jacques E. Dumont, and Carine Maenhaut. 1997. 'Characterization of a Phosphoprotein Whose MRNA Is Regulated by the Mitogenic Pathways in Dog Thyroid Cells'. European Journal of Biochemistry 248(3):660–68.

Wise, Jo Ann, and Hua Lou. 2021. 'Messenger RNA Processing in Eukaryotes'. in Reference Module in Life Sciences.

Wolfe, Cecily J., Isaac S. Kohane, and Atul J. Butte. 2005. 'Systematic Survey Reveals General Applicability of "Guilt-by-Association" within Gene Coexpression Networks'. BMC Bioinformatics 6:1–10.

Wu, Yadong, Feng Liu, Siyang Luo, Xinhai Yin, Dengqi He, Jianguo Liu, Zhaohui Yue, and Jukun Song. 2019. 'Co-Expression of Key Gene Modules and Pathways of Human Breast Cancer Cell Lines'. Bioscience Reports 39(7):1–18.

Xia, Zheng, Lawrence A. Donehower, Thomas A. Cooper, Joel R. Neilson, David A. Wheeler, Eric J. Wagner, and Wei Li. 2014. 'Dynamic Analyses of Alternative Polyadenylation from RNA-Seq Reveal a 3'2-UTR Landscape across Seven Tumour Types'. Nature Communications.

Xiaohong, Wang, Zhao Jun, Guo Hongmei, and Qinqin Fan. 2019. 'CFLAR Is a Critical Regulator of Cerebral Ischaemia-Reperfusion Injury through Regulating Inflammation and Endoplasmic Reticulum (ER) Stress'. Biomedicine and Pharmacotherapy.

Xu, Wenlong, Anthony San Lucas, Zixing Wang, and Yin Liu. 2014. 'Identifying MicroRNA Targets in Different Gene Regions'. BMC Bioinformatics 15(Suppl 7):6–8.

Yang, Wanling, Nan Shen, Dong Qing Ye, Qiji Liu, Yan Zhang, Xiao Xia Qian, Nattiya Hirankarn, Dingge Ying, Hai Feng Pan, Chi Chiu Mok, Tak Mao Chan, Raymond Woon Sing Wong, Ka Wing Lee, Mo Yin Mok, Sik Nin Wong, Alexander Moon Ho Leung, Xiang Pei Li, Yingyos Avihingsanon, Chun Ming Wong, Tsz Leung Lee, Marco Hok Kung Ho, Pamela Pui Wah Lee, Yuk Kwan Chang, Philip H. Li, Ruo Jie Li, Lu Zhang, Wilfred Hing Sang Wong, Irene Oi Lin Ng, Chak Sing Lau, Pak Chung Sham, and Yu Lung Lau. 2010. 'Genome-Wide Association Study in Asian Populations Identifies Variants in ETS1 and WDFY4 Associated with Systemic Lupus Erythematosus'. PLoS Genetics.

Yang, Yang, Wen Gao, Xi Ding, Wen Xu, Di Liu, Bo Su, and Yifeng Sun. 2017. 'Variations within 3'-UTR of MDM4 Gene Contribute to Clinical Outcomes of Advanced Non-Small Cell Lung Cancer Patients Following Platinum-Based Chemotherapy'. Oncotarget 8(10):16313–24.

Ye, Ying, Guangdong Wang, Guoyu Wang, Juhua Zhuang, Saifei He, Yanan Song, Jing Ni, Wei Xia, and Jiening Wang. 2017. 'The Oncogenic Role of Tribbles 1 in Hepatocellular Carcinoma Is Mediated by a Feedback Loop Involving MicroRNA-23a and P53'. Frontiers in Physiology.

Yip, Andy M., and Steve Horvath. 2007. 'Gene Network Interconnectedness and the Generalized Topological Overlap Measure'. BMC Bioinformatics.

Yokoyama, Takashi, Yohei Kanno, Yukari Yamazaki, Tomoko Takahara, Satoshi Miyata, and Takuro Nakamura. 2010. 'Trib1 Links the MEK1/ERK Pathway in Myeloid Leukemogenesis'. Blood 116(15):2768–75.

Yoshida, Akihiro, Jun Ya Kato, Ikuko Nakamae, and Noriko Yoneda-Kato. 2013. 'COP1 Targets C/EBPα for Degradation and Induces Acute Myeloid Leukemia via Trib1'. Blood.

Yoshino, Seiko, Takashi Yokoyama, Yoshitaka Sunami, Tomoko Takahara, Aya Nakamura, Yukari Yamazaki, and Shuichi Tsutsumi. 2021. 'Trib1 Promotes Acute Myeloid Leukemia Progression by Modulating the Transcriptional Programs of Hoxa9'. 137(1):75–88.

Yoshino, Seiko, Takashi Yokoyama, Yoshitaka Sunami, Tomoko Takahara, Aya Nakamura, Yukari Yamazaki, Shuichi Tsutsumi, Hiroyuki Aburatani, and Takuro Nakamura. 2020. 'Trib1 Promotes Acute Myeloid Leukemia Progression by Modulating the Transcriptional Programs of Hoxa9'. Blood.

Young, Matthew D., Matthew J. Wakefield, Gordon K. Smyth, and Alicia Oshlack. 2010. 'Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias'. Genome Biology 11(2).

Yu, Hui, Kang Tu, Yi Jie Wang, Jun Zhe Mao, Lu Xie, Yuan Yuan Li, and Yi Xue Li. 2012. 'Combinatorial Network of Transcriptional Regulation and MicroRNA Regulation in Human Cancer'. BMC Systems Biology 6.

Yuan, Lin, Haiyan Chu, Meilin Wang, Xiaojian Gu, Danni Shi, Ma Lan, Dongyan Zhong, Mulong Du, Pu Li, Na Tong, Guangbo Fu, Chao Qin, Changjun Yin, and Zhengdong Zhang. 2013. 'Genetic Variation in DROSHA 3'UTR Regulated by Hsa-MiR-27b Is Associated with Bladder Cancer Risk'. PLoS ONE.

Yuan, Ye, and Joanne B. Weidhaas. 2019. 'Functional MicroRNA Binding Site Variants'. Molecular Oncology.

Zhang, Bin, and Steve Horvath. 2005. 'A General Framework for Weighted Gene Co-Expression Network Analysis'. Statistical Applications in Genetics and Molecular Biology 4(1).

Zhang, Feng, Yulan Lu, Sijia Yan, Qinghe Xing, and Weidong Tian. 2017. 'SPRINT: An SNP-Free Toolkit for Identifying RNA Editing Sites'. Bioinformatics.

Zhang, Ruixian, Bangpin Pan, Yi Li, and Xiaolan Li. 2019. 'SNP Rs4937333 in the MiRNA-5003-Binding Site of the ETS1 3'-UTR Decreases ETS1 Expression'. Frontiers in Genetics.

Zhang, Xuelian, Bin Zhang, Chenyang Zhang, Guibo Sun, and Xiaobo Sun. 2021. 'Current Progress in Delineating the Roles of Pseudokinase Trib1 in Controlling Human Diseases'. Journal of Cancer.

Zhang, Zhe, Guojun Zhang, Chuize Kong, Jianbin Bi, Daxin Gong, Xiuyue Yu, Du Shi, Bo Zhan, and Peng Ye. 2015. 'EIF2C, Dicer, and Drosha Are up-Regulated along Tumor Progression and Associated with Poor Prognosis in Bladder Carcinoma'. Tumor Biology.

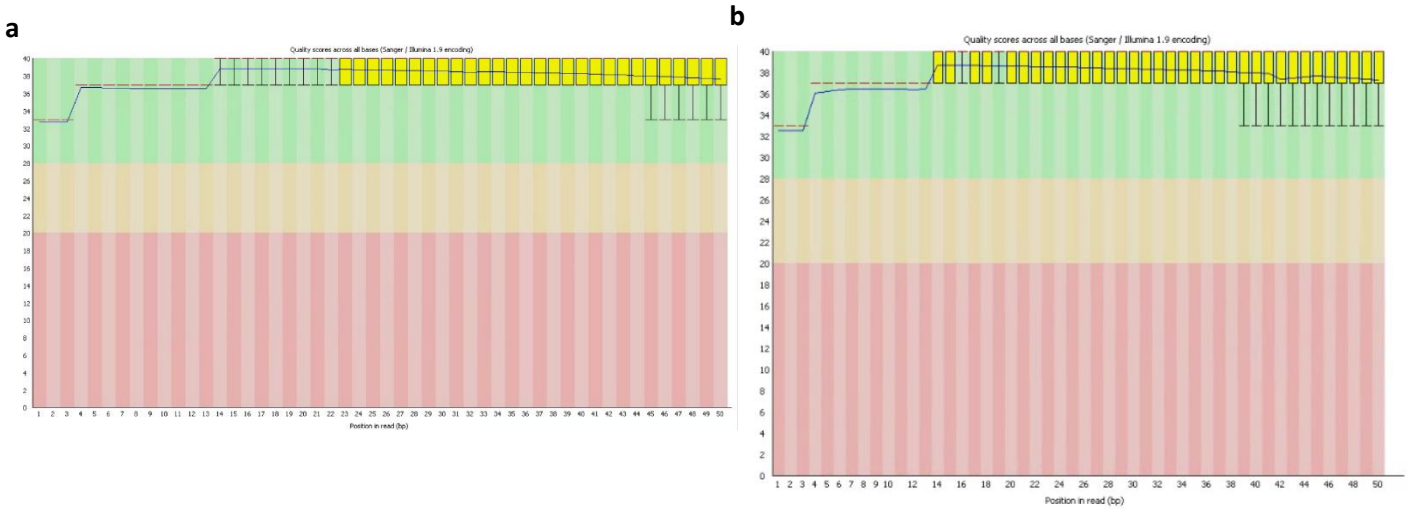
Ziegler-Heitbrock, Loems, Petronela Ancuta, Suzanne Crowe, Marc Dalod, Veronika Grau, Derek N. Hart, Pieter J. M. Leenen, Yong Jun Liu, Gordon MacPherson, Gwendalyn J. Randolph, Juergen Scherberich, Juergen Schmitz, Ken Shortman, Silvano Sozzani, Herbert Strobl, Marek Zembala, Jonathan M. Austyn, and Manfred B. Lutz. 2010. 'Nomenclature of Monocytes and Dendritic Cells in Blood'. Blood.

Zipeto, Maria Anna, Qingfei Jiang, Etienne Melese, and Catriona H. M. Jamieson. 2015. 'RNA Rewriting, Recoding, and Rewiring in Human Disease'. Trends in Molecular Medicine.

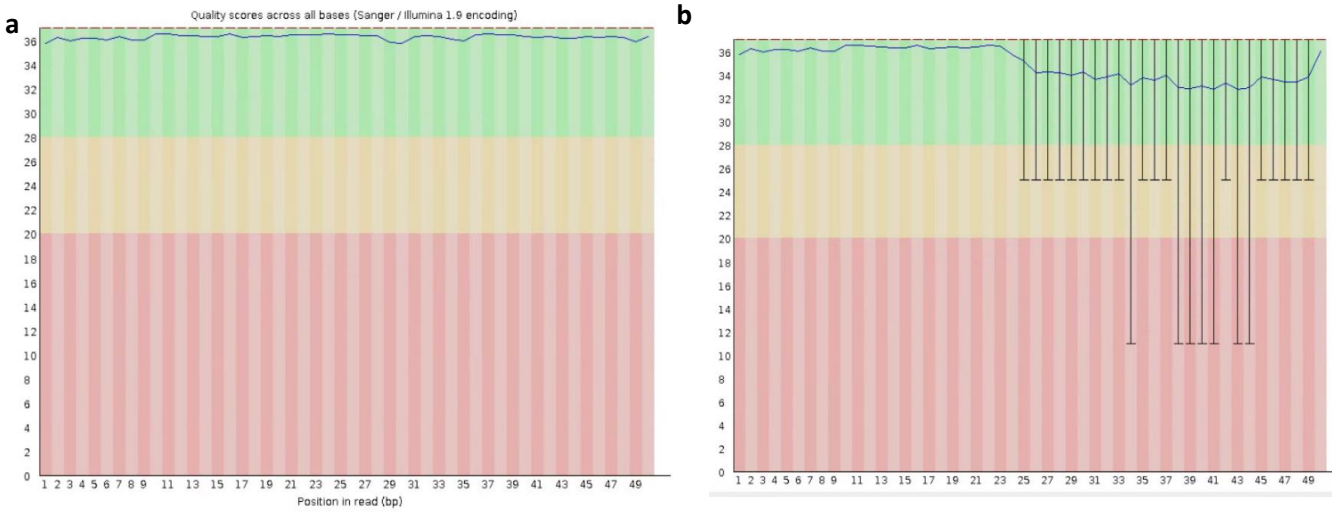
Appendix

Table A. 1- list of Software's/Tools and their version

Software/tool	Version
R	4.1
Python	3.1
Limma	3.48.3
EdgeR	3.34.1
DESeq2	1.32.0
GOSeq	1.46.0
BioMart	2.48.3
Keggrest	1.32.0
WGCNA	1.70.3
Ggplot2	3.3.5
gplots	3.1.1
CompleaxHeatmap	2.8.0
QuASAR	1
SPRINT	1
GATK	4.2
Picard	2.0.1
miRanda	3.3a
TargetScan	7.1
FastQC	3
Trimmomatic	0.38
Bedtools	2.28.0
Bcftools	1.14
Freebayes	1.35
Vcftools	0.1.13
MBASED	3.14
Samtools	1.14
RNAcentral	19
Procomp	3.6.2
tximport	4.1
DaPars	0.9.1
BWA	0.7.17
STAR	2.7.9a
Salmon	1.6.0



Supplementary figure 1. 1- FastQC box plots of quality scores per read position of macrophage RNA-seq data; a) Before adapter trimming, b) After adapter trimming



Supplementary figure 1. 2- Supplementary figure 1. 1- FastQC box plots of quality scores per read position of macrophages mi-RNA-seq data; a) Before adapter trimming, b) After adapter trimming

Table A. 2- Quality table summary of TRIB1 OE and control RNA-Seq data; Raw reads - total amount of reads of raw data, each four lines taken as one unit. For paired-end sequencing, it equals the amount of read1 and read2, Q30 - Base count of Phred value > 30) / (Total base count), GC content: (G & C base count) / (Total base count*)*

GFP_0_mins1	45095138	94.07	50.59
GFP_30_mins1	44472354	94.56	50.69
GFP_60_mins1	41591016	94.64	50.58
TRIB1_OE_0_mins1	42337430	94.63	50.35
TRIB1_OE_30_mins1	46249706	94.05	50.98
TRIB1_OE_60_mins1	47084062	94.81	50.74
GFP_0_mins2	44145900	94.72	50.38
GFP_30_mins2	43026548	94.45	51.99
GFP_60_mins2	57479166	94.73	51.96
TRIB1_OE_0_mins2	41941084	94.72	52.31
TRIB1_OE_30_mins2	48779126	94.44	52.2
TRIB1_OE_60_mins2	44614126	94.74	52.14
GFP_0_mins3	45652396	94.58	51.51
GFP_30_mins3	44209144	94.82	51.07
GFP_60_mins3	44829780	94.23	51.77
TRIB1_OE_0_mins3	40845944	94.7	51.69
TRIB1_OE_30_mins3	46091602	94.54	51.93
TRIB1_OE_60_mins3	44716110	94.74	51.96

Table A. 3- Quality table summary of miRNA OE in unstimulated macrophages RNA-seq data; Raw reads - total amount of reads of raw data, each four lines taken as one unit. For paired-end sequencing, it equals the amount of read1 and read2, Q30 - Base count of Phred value > 30) / (Total base count), GC content: (G & C base count) / (Total base count)

d5a	22776488	93.14	50.42
d5b	26134056	94.86	50.77
d5c	27155122	92.78	50.63
d5d	22566917	93.34	50.62
d7a	21827616	93	50.34
d7b	26869493	92.54	49.66
d7c	28342110	92.9	49.44
d7d	24411898	93.19	49.71
d12a	27706647	92.39	50.86
d12b	27768686	93.31	50.03
d12c	27073927	93.03	50.39
d12d	24034943	93.08	50.19
d14a	26708208	94.37	50.55
d14b	21870461	93	49.77
d14c	22143558	92.71	49.97
d14d	22625692	93.27	50.54
d16a	23165588	93.4	50.75
d16b	24611795	92.75	50.7
d16c	27877058	93.12	50.7
d16d	27727354	93.12	50.79
d17a	22425135	93.37	50.64
d17b	24008424	94.65	50.97
d17c	27833959	94.78	50.75
d17d	22313814	94.88	50.7

Table A. 4- Mapping percentage of small non-coding RNA seq in human polarised macrophages: Mun vs MLPS+IFNγ

Sample	Mapping percentage
Sample_A_M0	69%
Sample_A_M1	71.60%
sample_C_M1	70.67%
sample_C_M0	76.11%
Sample_D_M0	72.26%
Sample_D_M1	65.36%
Sample_E_M0	72%
Sample_E_M1	67%
sample_F_M0	71.48%
sample_F_M1	73%
sample_G_M0	70%
sample_G_M1	66.45%
Sample_H_M0	72.78%
Sample_H_M1	73%
Sample_I_M0	70.88%
Sample_I_M1	73.11%

Table A. 5- Mapping percentage RNA seq in human polarised macrophages: Mun vs MLPS+IFNγ

Sample	Mapping percentage
RHM5411	91%
RHM5412	87.75%
RHM5417	91.53%
RHM5418	90.66%
RHM5423	91.74%
RHM5424	91.32%
RHM5429	85.20%
RHM5430	89.40%
RHM5435	90.94%
RHM5436	88.82%
RHM5441	90.40%
RHM5442	87.43%
RHM5447	90.43%
RHM5448	88.78%
RHM5453	90.08%
RHM5454	90.36%

Table A. 6- Mapping percentage of RNA-seq on unpolarised MDMs transiently transfected with miRNA mimics/control

Samples	Mapping percentage
d1a	90
d1b	82
d1c	8
d1d	86
d2a	90
d2b	91
d2c	90
d2d	93
d3a	91
d3b	90
d3c	90
d3d	89
d4a	91
d4b	91
d4c	91
d4d	92
d4d	92
d5b	90
d5c	92
d5d	91
d6a	91
d6b	88
d6c	90
d6d	91

Table A. 7- Mapping percentage of RNAseq, performed on GFP control and TRIB1-overexpressed mRNA

Samples	Mapping percentage
GFP_0_mins	83
GFP_0_mins2	83
GFP_0_mins3	82
GFP_30_mins1	83
GFP_30_mins2	82
GFP_30_mins3	84
GFP_60_mins1	83
GFP_60_mins2	84
GFP_60_mins3	84
TRIB1_OE_0_mins1	82
TRIB1_OE_0_mins2	82
TRIB1_OE_0_mins3	83
TRIB1_OE_30_mins1	83
TRIB1_OE_30_mins2	83
TRIB1_OE_30_mins3	83
TRIB1_OE_60_mins1	83
TRIB1_OE_60_mins2	82
TRIB1_OE_60_mins3	82