Identification of differentially expressed lncRNAs in CD4+ T cell memory compartments

A Thesis

submitted to

Indian Institute of Science Education and Research Pune in partial fulfilment of the requirements for the BS-MS Dual Degree Programme

by

Fajar Pm

Indian Institute of Science Education and Research Pune

Dr. Homi Bhabha Road,

Pashan, Pune 411008, INDIA.

Date: April, 2024

Under the guidance of

Supervisor: Dr. Veena S. Patil

Scientist V, National Institute of Immunology, New Delhi

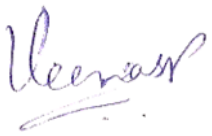
From June 2023 to Mar 2024

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled ‘identification of differentially expressed lncRNAs in CD4+ T cell memory compartments’ towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Fajar Pm at National Institute of Immunology, New Delhi under the supervision of Dr. Veena S. Patil, Scientist V, Department of infection and Immunity, during the academic year 2023-2024.

Dr. Veena S. Patil



Committee:

Dr. Veena S. Patil

Scientist V

National Institute of Immunology

Dr. Satyajit Rath

Professor Emeritus

Department of biology, IISER Pune

This thesis is dedicated to my family

Declaration

I hereby declare that the matter embodied in the report entitled “Identification of differentially expressed lncRNAs in CD4+ T cells memory compartments” are the results of the work carried out by me at the Department of Infection and Immunity, National Institute of Immunology, New Delhi, under the supervision of Dr. Veena S. Patil and the same has not been submitted elsewhere for any other degree

Fajar Pm

20191106

Date: 13/03/24

Table of contents

List of Figures

List of Tables

Abstract

Long non-coding RNAs (lncRNAs) are RNA molecules longer than 200 nucleotides that do not code for proteins but have been shown to perform regulatory roles in various cellular processes and in disease conditions. They have also been shown to be involved in the regulation of cellular processes, disease conditions, and lineage-specificity. Since the lncRNAs harbor immense regulatory potential at both transcriptional and post-transcriptional levels, identifying and molecular characterizing T cell memory subset-specific lncRNAs will add new dimensions to understanding T cell memory development and functioning. Hence, as part of my MS thesis, I analyzed the transcriptomic (RNA-Seq) and epigenomics (ATAC-Seq) data from six memory CD4+ T cell subsets to identify memory subset-specific lncRNAs. Comparative analysis of differentially expressed genes identified 42 and 36 upregulated lncRNAs in long-term (TCM, TSCM) and short-term effector memory subsets (TEM, T_{EMRA}-precursor, and T_{EMRA}-effector), respectively. Further, to identify the potential cis-regulatory role of these lncRNAs, I identified differentially expressed protein-coding genes located upstream and downstream of these lncRNAs. This analysis has identified 16 such lncRNA-protein-coding genes pairs. Further functional characterization of these lncRNA-protein coding gene pairs for the co-regulation can yield interesting insights into the T cell memory development and commitment.

Acknowledgments

First and foremost, I thank my supervisor, Dr. Veena S. Patil, for her constant support, invaluable advice, and helpful discussions. Her excitement about my work and her willingness to try out new creative directions of questioning have encouraged and inspired me throughout the dissertation. I would also like to thank Raunak Kar, a senior Ph.D. student at the Immunogenomics lab, for his valuable guidance, enlightening discussions, and constant support throughout. He patiently helped me to make things happen. I also thank other lab members, Shreya and Kirti, for making me comfortable in the lab, and Nandan sir and Sarojini ma'am for the technical assistance.

I also want to express my gratitude to Dr. Satyajit Rath for his valuable guidance and suggestions. His scientific thoughts and teaching really inspired me to take a research career in immunology. Finally, I would like to express my gratitude to my close friends and my amazingly supportive family, who always believed in me even when I did not. Without their tremendous support and encouragement over the past year, it would be impossible for me to complete this study. Finally, I would like to thank IISER Pune for the studentship that allowed me to conduct this thesis.

Contributions

Contributor name	Contributor Role
Dr. Veena S. Patil	Conceptualization ideas
Raunak kar, Shreya Sinha	Methodology
Nil	Software
Dr. Veena S. Patil	Validation
Nil	Formal analysis
Nil	Investigation
Dr. Veena S. Patil	Resources
Nil	Data Curation
Nil	Writing – original draft preparation
Raunak kar	Writing – review and editing
Dr. Veena S. Patil	Visualization
Dr. Veena S. Patil	Supervision
Dr. Veena S. Patil	Project administration
Dr. Veena S. Patil	Funding acquisition

Introduction

One of the most important facets of the adaptive arm of the immune system is immunological memory. Once the immune system encounters an antigen, the naïve T cells, upon recognizing the particular antigen, give rise to short-lived effectors and long-term memory T cells. The effector T cells, after successful clearance of said pathogen, die, while a pool of memory cells stays behind to give long-lasting protection to the individual upon subsequent encounters with the same antigen. The memory cells generated in response to infection are heterogeneous and can be further classified majorly as long-term precursor like and short-term effector like memory. Based on their longevity and effector profile, conventionally, memory is broadly classified into stem cell like memory (TSCM), central memory (TCM), effector memory (TEM), and short-lived effector memory expressing CD45RA (TEMRA) (Mahnke et al., 2013) (fig 1).

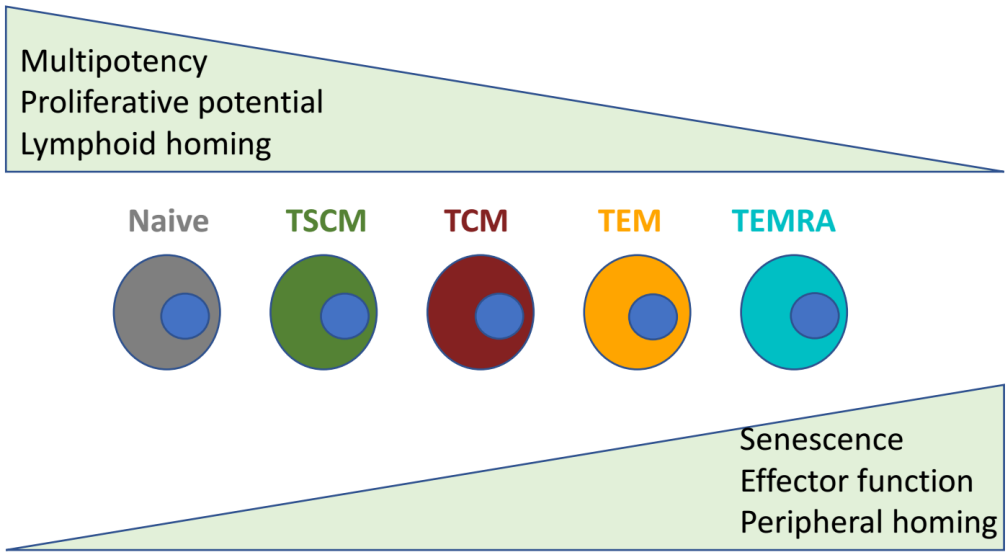


Figure 1. T cell memory compartments. Correlation of phenotypes with memory compartments are pictorially represented.

T memory stem cells (TSCMs) are memory T cells that possess an increased potential to self-renew and has the multipotent ability to give rise to central memory, effector memory, and effector T cells. TSCM exhibits a comparable expression pattern of surface receptors to naïve (like CCR7 & CD45RA). However, they can be separated from naive by the expression of two markers, CD95 and IL-2Rß (Gattinoni et al., 2011).

TCMs are primarily limited to lymphoid organs and the blood. They are functionally distinguished by their higher proliferative potential after antigen reencounter and the ability to seed peripheral tissues with additional effector cells following activation. In contrast, TEMs are found in the circulation and have the potential to travel via peripheral organs. They exhibit quick effector activities but have limited proliferative potential, serving as first responders in cases of reinfection (Sallusto et al., 1999).

TEMRA, which is TEM expressing CD45RA, is a compartment within TEM and is highly associated with cytotoxic function (Sallusto et al., 2004). TEMRA compartments were initially attributed to CD8+ T cells, where they were associated with direct killing of virus infected or transformed cells. However, studies have shown despite being helper cells, CD4+ T cells also have the ability to become cytotoxic (CD4-CTL) in nature (Juno et al., 2017). Later, people have shown that this CD4-CTL is enriched in the CD4 TEMRA compartment is heterogenous and can be further classified into TEMRA precursors and TEMRA effectors based on the surface marker CD127 (receptor for IL-7), a long-term survival marker of naïve and long-term memory subset. TEMRA cells expressing CD127 are TEMRA precursors, and those that do not are TEMRA effectors (Patil et al., 2018).

Large-scale changes at both transcriptome and epigenetic levels happen when activated T cells differentiate into a heterogeneous pool of memory cells (Araki et al., 2009; Youngblood et al., 2013; Chen et al., 2018). Several studies have been done to investigate the role of differentially expressed protein-coding genes responsible for creating heterogeneity within the T cell memory (Kaech et al., 2002; Araki et al., 2009; Eshima et al., 2018; Omilusik et al., 2015; Kaech and Cui, 2012). Therefore, the role of protein-coding genes in differentiating T cells into a heterogeneous memory pool is widely studied. However, the change in transcriptome upon differentiation is not limited to protein-coding genes; recent studies, including ours, have shown that non-coding genes also show differential expression (Hudson et al., 2019; Spurlock et al., 2017; Zhang et al., 2014). This implies that non-coding genes may also play a role in the differentiation of naive T cells into different memory compartments and may help a better understanding of T cell biology.

The non-coding part of the genome is comprised of different types of RNAs. Among these, RNAs that are longer than 200 nucleotide base pairs are long non-coding RNAs (lncRNAs), which represent the major portion of the non-coding genome in humans GENCODE ((https:// www.gencodegenes.org). Even though lncRNAs don't code for any protein, they share features with mRNA. Like, they are spliced and polyadenylated, even though there are exceptions (Mattick et al., 2023). lncRNAs are often defined by their relationships with neighbouring protein-coding genes. As a result, with respect to protein-coding genes, they are classified as intergenic, intronic, or antisense (Mattick et al., 2023). One of the striking features of lncRNAs is that they show more restricted expression patterns than mRNAs (Derrien et al., 2012; Gloss and Dinger, 2016). Because of their low expression pattern and lack of coding potential, lncRNAs

were initially thought to be transcriptional noise. But later, advancements in technologies like RNA-seq revealed the differential expression of many lncRNAs in different cell types, and it began to be a hot research topic. Then, there was a surge in the study of lncRNAs, which led to the functional characterization of many lncRNAs and showed that lncRNAs were capable of performing a wide array of functions, such as organising nuclear domains, regulating proteins or RNA molecules, and controlling transcription either in cis or trans (Ulitsky and Bartel, 2013**)** (fig 2).

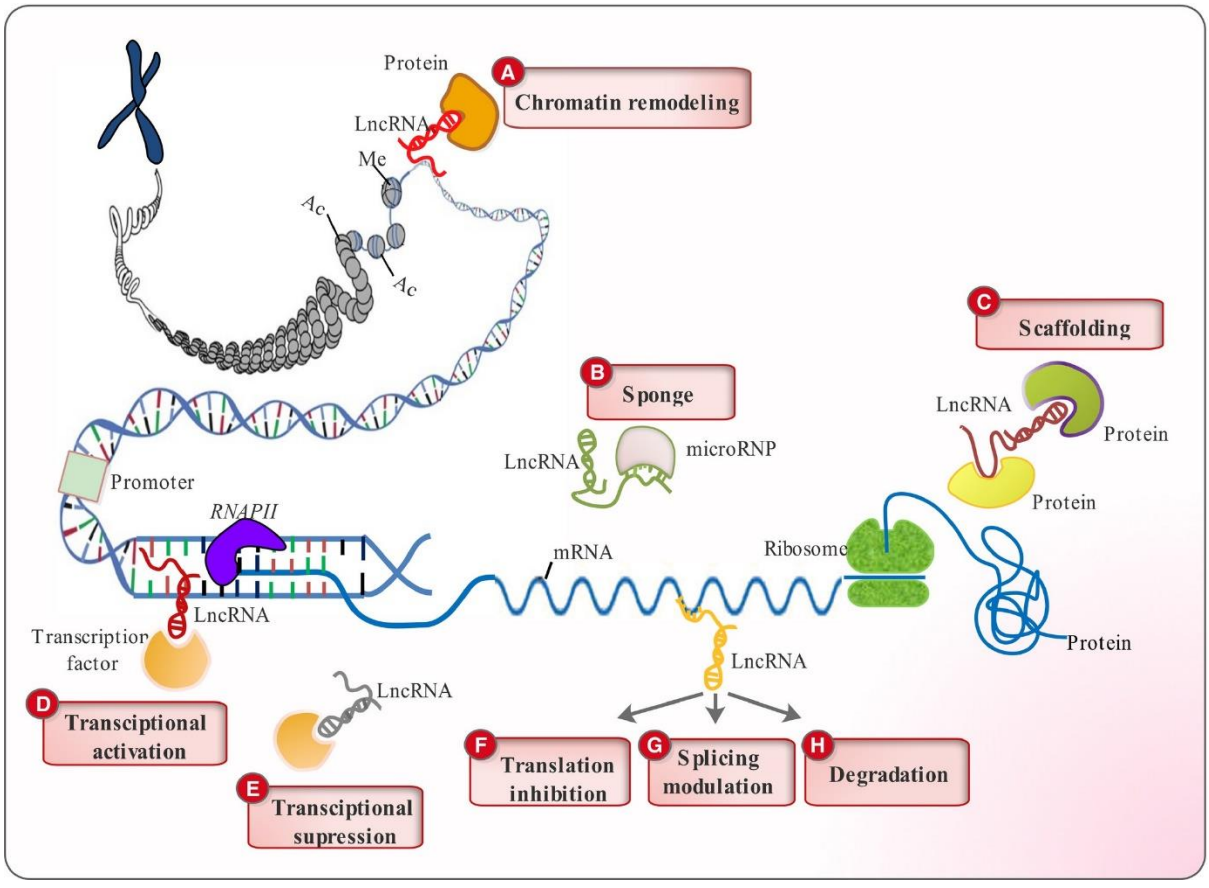


Figure 2. (Salehi et al., 2017) Classification of lncRNA functions. (A) Chromatin organizational patterns can be altered by lncRNAs through the recruitment of distinct protein components of the chromatin remodelling complex. (B) By base pairing with their complementary miRNAs and mitigating their effects, they can function as "sponges.". (C) LncRNAs can act as scaffolds by giving proteins that work in the same biological pathway docking sites. (D) By directing transcription factors to their promoters, they stimulate the transcription of specific genes. (E) By enclosing transcription factors and preventing them from connecting to their promoters, lncRNAs have the ability to suppress transcription. Through base pairing, (F) translation inhibition, (G) splicing pattern modification, and (H) exposure to degradative pathways, they can modify the functioning of mRNA.

lncRNAs are known for their regulatory roles in various cellular processes and in disease conditions (Patil et al., 2014; Li et al., 2021), and have been reported to regulate protein-coding genes in cis or trans (Chang et al., 2020; Spurlock III et al., 2015; Zhang et al., 2019**).** They have also been studied in the context of T cells. lncRNAs have been shown to regulate key protein-coding genes associated with T cells (Spurlock III et al., 2015; Petermann et al., 2019). lncRNA was also reported to be involved in specifying the lineage in T cell memory compartments (Hudson et al., 2019; Spurlock et al., 2017; Ranzani et al., 2015). Since lncRNAs harbour immense regulatory potential at both transcriptional and post-transcriptional levels, identifying subset-specific lncRNAs will add new dimensions to understanding T cell's development and functioning.

Several studies have been done in this regard to understand the role of lncRNAs in T cell memory development, but the studies were limited to characterizing lncRNAs in conventional memory subsets like TCM, TEM, and helper subsets in the case of CD4+ T cells (Ranzani et al., 2015; Spurlock et al., 2017**)**. A more comprehensive picture of lncRNA regulation in CD4+ T cell memory development remains unclear due to the fact that the memory compartments of CD4+ T cells are highly heterogeneous and the investigations were restricted to a subset of memory compartments. Hence, this project focuses on understanding the role of lncRNAs in the differentiation of CD4+ T cell memory development at a higher resolution by further dividing long-term memory compartments into TSCM and TCM, and effector-memory compartments into TEM, TEMRA-P, and TEMRA-E. The specific objectives of my thesis project are as follows.

Specific objectives

1) Identification of differentially expressed lncRNAs in CD4+ T cell memory compartments including TSCM, TCM, TEM, TEMRA-P & TEMRA-E.

2) Selection of potential cis-regulatory lncRNAs.

Materials and methods

Bulk RNA-seq data of CD4+ T cell memory compartments

The bulk RNA-seq data of CD4+ T cell memory compartments (TSCM, TCM, TEM, TEMRA-P and TEMRA-E), including naive (from 10 donors) done by Kar et al (unpublished), was used. Briefly, the bulk RNA-seq was performed using the smart-seq2 method (Picelli et al., 2014). GRCh38 (hg38) was used as the reference genome. The counts were normalized to Transcripts per million (TPM). The normalized count matrix was used to check the expression level of protein-coding and lncRNA genes across the memory compartments (TSCM, TCM, TEM, TEMRA-P and TEMRA-E), including naive. The normalized counts were also used to calculate the correlation between the selected lncRNAs and their corresponding cis gene.

Differential expression analysis of bulk RNA-seq data

Differential expression (DE) analysis on the bulk RNA-seq data was done by Kar et al (unpublished). Briefly, the DEseq2 package takes RNA-seq raw count as input to do differential expression analysis of genes between two samples or conditions (here different memory compartments). Data from DE analysis was used to identify differentially expressed lncRNAs between the memory compartments where the information about the log2Fold change and the corresponding significance (Padj value) are available. This data was analysed to identify differentially expressed lncRNAs and used to make potential lncRNA;mRNA pairs

Principal Component Analysis (PCA)

Qlucore Omics Explorer 3.8 software package was used to perform PCA of the transcriptomes. Normalized gene count matrix generated by bulk RNA-seq of CD4+ T cell memory compartments mentioned above was used. The normalized gene count matrix was uploaded to Qlucore Omics Explorer. Cell types were annotated as samples and genes as variables. 1000-most variable protein-coding and lncRNA genes were identified by software and are used for PCA. The corresponding heatmaps were also generated using the software.

Weighted Gene Correlation Network Analysis (WGCNA).

The WGCNA was done by Kar et al (unpublished). A weighted gene network was created by using the WGCNA package's standard process using the depth-normalized expression data obtained from the RNA-seq studies as input (Zhang et al., 2005). WGCNA on RNA-seq data resulted in 55 modules. Each module represents a set of genes that are co-expressed across the samples (here, memory compartments). Each module was correlated with two traits: cell type and memory type. Modules with significant correlation with traits were re-analyzed using Qlucore Omics Explorer software to figure out the modules of interest (modules having the selected lncRNAs).

Identification of common lncRNAs of interest using Venn diagram

Lists of lncRNAs differentially upregulated in individual memory compartments were made. Then, to identify common lncRNAs between two or more compartments of interest, a software called Venny 2.1 was used. To identify common lncRNAs between two or more compartment, list of lncRNAs differentially upregulated corresponding to each compartment was uploaded in Venny 2.1. Based on the input, Venny 2.1 will give Venn diagram showing the common lncRNAs. It will also give the list of common lncRNAs of interest.

Correlation analysis of lncRNA:mRNA pairs

Spearman correlation between the lncRNA:mRNA pairs selected was performed using GraphPad Prism 8.0.2. All the correlations were statistically significant (P (two-tailed) < 0.0001)

ATAC-seq data and tracks generated in the UCSC genome browser

The tracks in the UCSC genome browser and ATAC-seq data were generated by Kar et al (unpublished). Briefly, ATAC-seq sequences were quality filtered and adaptor trimmed using trim_galore, and trimmed reads were mapped to the GRCh38 (hg38) human reference sequence using Bowtie2. De-duplicated reads (using picard markduplicates) were filtered for reads mapping to ENCODE blacklisted region, mitochondrial genome and low-quality reads (bedtools and samtools). Reads passing all quality checks were called for peaks using MACS2. Peaks were visualized using UCSC Genome Browser after creating manual tracks for each sample using BIGWIG file format which

were further generated using bedgraphtobigwig. To check the openness of chromatin and to visually represent the RNA expression level of the lncRNA-mRNA pairs selected, tracks generated in the UCSC genome browser was used. The tracks for each lncRNA:mRNA pair was analysed individually to get the status of openness of the chromatin at transcription start site (TSS) and at exon level.

Results

Transcriptome analysis of CD4+ T cell memory subsets revealed differential expression of lncRNAs.

T cell memory has been shown to be highly heterogeneous and can be classified into different compartments based on their longevity, proliferative potential, and effector status, namely naive (TN), stem cell-like memory (TSCM), central memory (TCM), effector memory (TEM) and effector memory expressing CD45RA (TEMRA). This heterogeneity is largely defined by the expression of memory-specific molecules, including transcription factors, surface receptors, and effector molecules. To check whether the lncRNAs are also differentially expressed across the memory compartments, principal component analysis (PCA) on the expression of the 1000 most variable protein-coding genes and lncRNAs were compared. As expected, the protein-coding genes clustered the different memory compartments into two (PC1, 32% of variance), where long-lived TSCM and TCM clustered with naive (termed as long-term memory cluster) and short-lived effector TEM, TEMRA-P and TEMRA-E clustered together (termed as effector memory cluster) (Fig 3a). Interestingly, when lncRNAs were considered, they also clustered the same way as protein-coding genes did (PC1, 20% of variance), clustered into long-term and effector memory, revealing that lncRNAs were highly differentially expressed across memory compartments (Fig 3b). Heatmap of the genes dictating the principal component (PC) of the protein-coding (Fig 3c) and lncRNAs (Fig 3d) provided further verify the observation.

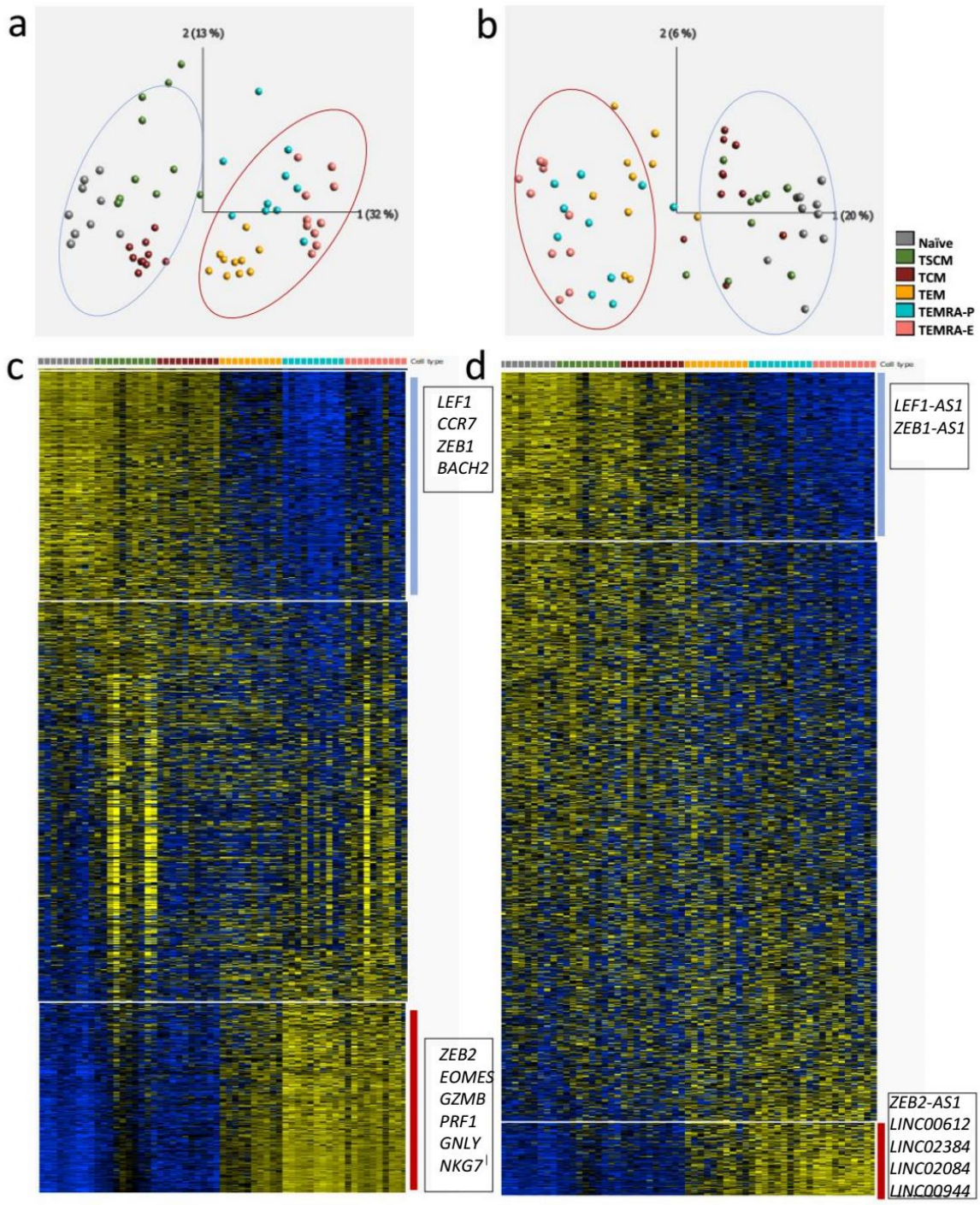


Figure 3. Differential expression of lncRNAs across CD4+ T cell memory compartments. Plot of first two principal components (PC) (PC1 and PC2 shown) of 1000 most variable protein-coding genes (a) and lncRNAs (b) from different memory compartments (colour coordinated) from 10 donors. The cluster marked with red is effector memory cluster and the other cluster marked with light blue is the long-term memory cluster. Percentage of variation explained by each PC is shown in parentheses. Heatmap showing the row-wise normalized z-scores of 1000 most variable protein-coding (c) and lncRNAs (d) across the CD4+ T cell memory compartments. Some of the genes from the marked region are mentioned in box. Samples (cell types) are arranged in the order of their effector profile (naïve < TSCM < TCM < TEM < TEMRA-P < TEMRA-E from left to right as highlighted and variables (genes) are arranged based on their corresponding PC1 of PCA.

Long-term memory and effector-memory specific lncRNAs

The PCA of lncRNAs clustered the CD4+ T cell memory compartments into long-term memory and effector memory compartments. So, to identify the long-term and effector memory specific lncRNAs primarily responsible for this clustering, differential expression (DE) analysis done by pairwise comparison of memory compartments (TN vs TSCM, TN vs TCM, etc. of all possible combinations) were used. For a lncRNA to be considered as upregulated, cut off of log2 fold change >/= 0.5 & Padj </= 0.05 were used.

Long-term memory specific lncRNAs: To identify lncRNAs upregulated in naive compared to effector memory, an intersection from the paired analysis of naive vs TEM, naive vs TEMRA-P and naïve vs TEMRA-E were taken (228 lncRNAs, Fig 4a). A similar analysis was done individually for TSCM (83 lncRNAs) and TCM (56 lncRNAs). Further, to narrow down lncRNAs specific to long-term memory T cells (TN, TSCM and TCM), the intersection of the individual lists was taken (42 lncRNAs, Fig 4b).

Effector memory-specific lncRNAs: A similar approach identified 36 lncRNAs commonly upregulated in TEM, TEMRA Pre, and TEMRA Eff compared to TN, TSCM & TCM and were referred to as effector-memory lncRNAs (fig 4c). The expression pattern of the memory specific lncRNAs clearly shows the segregation of lncRNAs selected into long-term and effector memory specific (fig 4d).

Long-term specific and effector memory specific lncRNAs identified in this analysis could serve as potential modulators, ensuring the longevity and effector-ness of the respective class of memory.

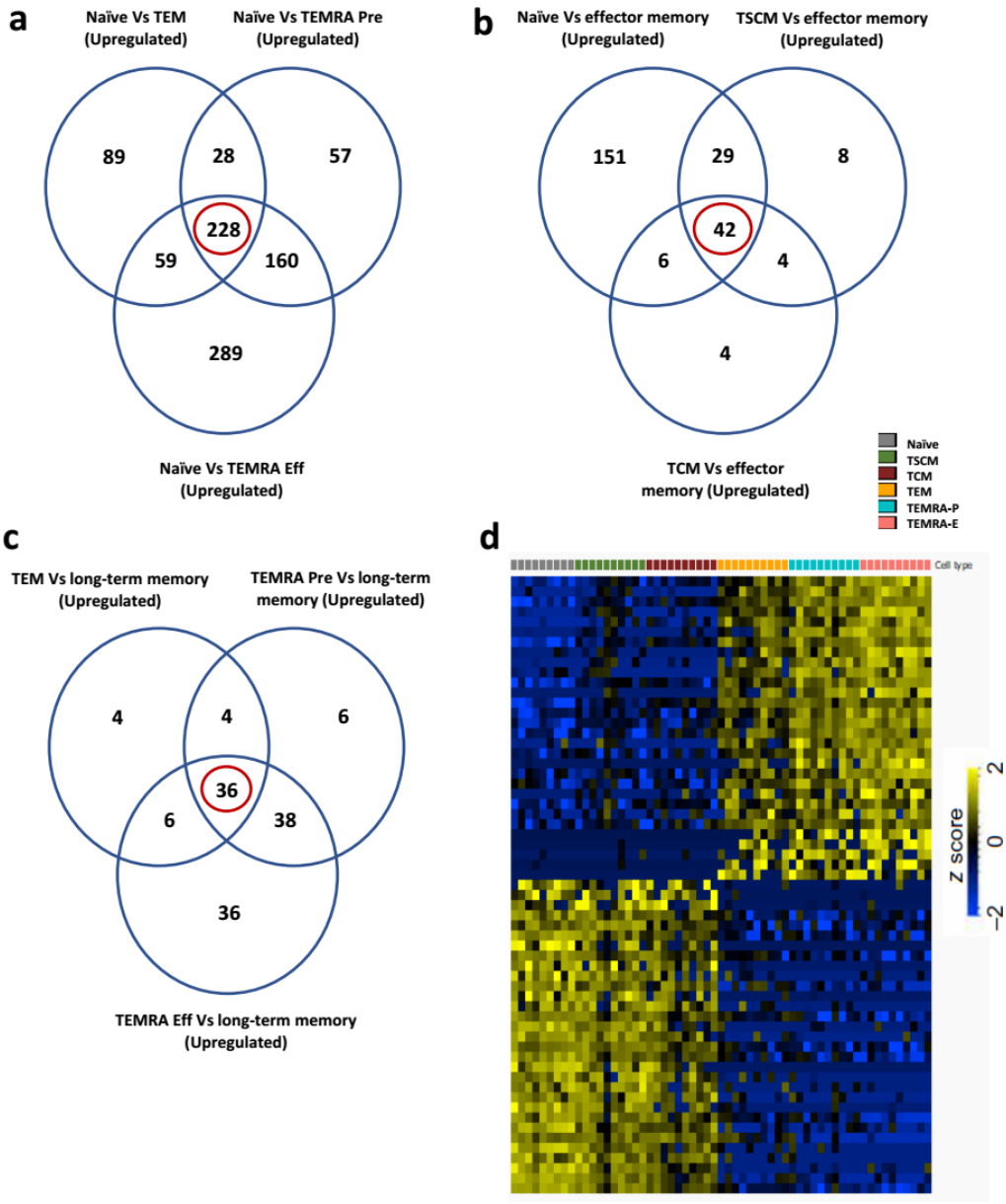


Figure 4. identification of long-term and effector memory lncRNAs. **a)** Venn diagram showing the identification of lncRNAs differentially upregulated in naïve compared to effector memory compartment (TEM, TEMRA Pre & TEMRA Eff. **b)** Venn diagram showing the lncRNAs commonly upregulated in naïve, TSCM and TCM compared to effector memory compartments and the intersection of these three as the long-term memory lncRNAs. **c)** Venn diagram showing the lncRNAs commonly upregulated in TEM, TEMRA Pre and TEMRA eff compared to long-term memory compartments and the intersection of these three as considered as the long-term memory lncRNAs. **d)** Heatmap showing the row-wise normalized z-score of the long-term memory and effector memory lncRNAs across CD4+ T cell memory compartments. Samples (cell types) are arranged based on their effector profile and variables (genes) are arranged based on the PC1 of PCA

Selection of lncRNAs for functional characterization

LncRNAs are shown to regulate protein-coding gene expression by acting in cis or trans (Chang et al., 2020; Spurlock III et al., 2015; Zhang et al., 2019). For the current analysis, I focussed on identifying cis-regulatory lncRNAs. To explore how lncRNAs function in T-cell memory, I looked closely into the memory-specific lncRNAs previously identified. This analysis revealed that some of the lncRNAs were cis to protein-coding genes studied in T cells, and were found to be co-expressed (for example, LEF1-AS1 identified as a long-term memory specific lncRNA is antisense to LEF1, which is known for its role in long-term memory). Conventionally, CD4+ T cells are known for their helper phenotype, where they provide help to other immune cells like B cells, monocytes, and CD8+T cells, while CD8+ T cells are known for their cytotoxic phenotype. Despite this classical dichotomy, CD4-TEMRA have been shown to exhibit cytotoxicity, as shown in the case of several viral infections and vaccination strategies (Weiskopf et al., 2015; Cheroutre et al., 2013; Watson et al., 2016), and have been shown to be correlate with protective phenotype. To investigate if lncRNAs have a role to play in defining the effector phenotype of these CD4 T cells, I first looked into effector memory specific lncRNAs identified above. I further grouped these effector memory lncRNAs into all effector memory (TEM, TEMRA-P and TEMRA-E) and ones specifically expressed by the TEMRA compartment. To be more stringent, I put the cutoff as log2foldchange >= 1. Next, using the UCSC genome browser, protein-coding genes coming upstream and downstream of selected lncRNAs were identified. Then, lncRNA-mRNA pairs were formed if the

protein-coding genes were also differentially expressed, like their corresponding lncRNAs in the respective memory compartment (fig 5).

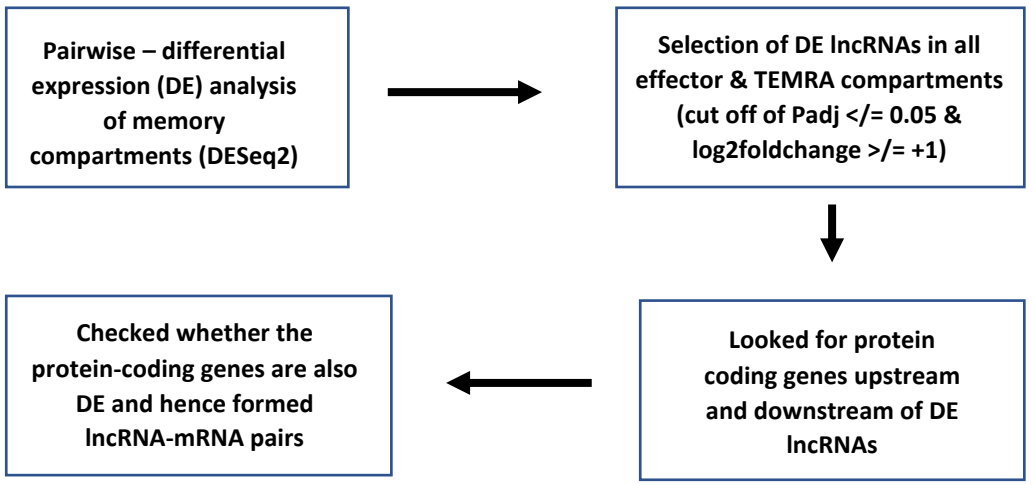


Figure 5. Flowchart showing the selection criteria of lncRNAs for functional characterization

Following the selection criteria mentioned above, 15 lncRNA-mRNA pairs were formed from the effector memory compartment and 1 lncRNA-mRNA pair from the TEMRA compartment (table 1). To narrow it down further, from the lncRNA-mRNA pairs formed, lncRNAs paired with protein-coding genes already studied and shown to be associated with effector T-cells were considered for functional characterization. ZEB2-AS1 and ZEB2-AS are two additional lncRNAs (cis to ZEB2, a transcription factor associated with effector memory) considered which are not coming from the above list. ZEB2-AS1 and ZEB2-AS are upregulated in the effector memory compartment but compared to other lncRNAs selected, they have a significant expression in TSCM (multipotent ability to derive TCM, TEM and effector T cells (Gattinoni, L. et al., 2011)) but not in naïve and TCM similar to ZEB2. This made them an interesting candidate.

lncRNAs	Cis genes	Compartment
KLRK1-AS1	KLRK1	TEMRA
LINC00612	KLRG1	All effector memory
LINC02384	IFN-γ	All effector memory
LINC02084	EOMES	All effector memory
FAM53B-AS1	FAM53B	All effector memory
MSC-AS1	MSC	All effector memory
A2M-AS1	PZP	All effector memory
DENND3-AS1	DENND3	All effector memory
B3GAT1-DT	B3GAT1	All effector memory
ENSG00000257438	NUAK1	All effector memory
ENSG00000271314	PTCH1	All effector memory
ENSG00000271387	C1orf21	All effector memory
ENSG00000234389	IL18RAP	All effector memory
ENSG00000235192	GALNT3	All effector memory
ENSG00000276241	CCL4	All effector memory
ENSG00000286942	NMUR1	All effector memory

Table 1. List of lncRNA-mRNA pairs formed following the selection criteria. lncRNAs which doesn't have an external gene name is indicated by their ensemble ID. The first pair is TEMRA specific (TEMRA-P & TEMRA-E) and the rest are from all effector memory compartment (TEM & TEMRA-P & TEMRA-E).

A parallel approach was taken to confirm that the lncRNAs considered for functional characterization (Table 3) were effector-memory specific. Weighted gene correlation network analysis (WGCNA) performed on the bulk RNA-seq of CD4+T cell memory compartments (mentioned previously) along with a well-characterized effector immune cell (CD8+ TEMRA compartment, (kar et al. unpublished) was used for this. 55 modules were identified in the dataset with correlated gene signatures, both positive and negative correlations. Traits of interest, such as memory type and cell type were given a numerical value to ascertain significance with the modules. Donor type was also used as a trait to discard modules which showed correlation significance with donors. Modules of

significant correlation with trait of interest were narrowed down to 19 and they were associated with early long-term memory specific, long-term memory specific, effector-memory specific and late effector memory specific signatures.

Detailed analysis of the modules revealed that the lncRNAs selected were coming in three modules (which were included in the 19 modules mentioned): yellow (4 lncRNAs), black (1 lncRNA), and darkorange2 (1 lncRNA) (fig 6). All three modules were enriched with effector-memory specific genes, and they were positively correlated with effector-memory type and late memory cell type trait (TEM, TEMRA-P, TEMRA-E & CD8 TEMRA) (Table 2). Therefore, the lncRNAs shortlisted were confirmed to be effector memory specific. This is further supported by the evidence that most of the lncRNAs shortlisted here are expressed by other well-characterized effector immune cells (NK cells and effector CD8+ T cell) DICE (https://dice-database.org/) (Hudson, W. H. et al., 2019). Therefore, I selected those six effector-memory specific lncRNAs for functional characterization (table 3). Taken together, with the help of differential expression analysis (CD4+ T cell memory compartments), data from WGCNA (by considering CD8+ TEMRA with CD4+ T cell memory compartments) and studies showing the expression of selected lncRNAs by well-characterized effector-immune cells DICE, I was able to show that the selected lncRNAs are effector immune-memory specific.

Module	Module-trait correlation (Memory type)	Module-trait correlation (Cell type)
Yellow	0.83 ($2*e^{-18}$)	0.75 ($2*e^{-13}$)
Black	0.71 ($1*e^{-11}$)	0.63 ($7*e^{-9}$)
Darkorange2	0.67 ($4*e^{-10}$)	0.39 ($1*e^{-3}$)

Table 2. The modules-trait correlation of the modules of interest. The Pearson correlations calculated are statistically significant with p values in parentheses.

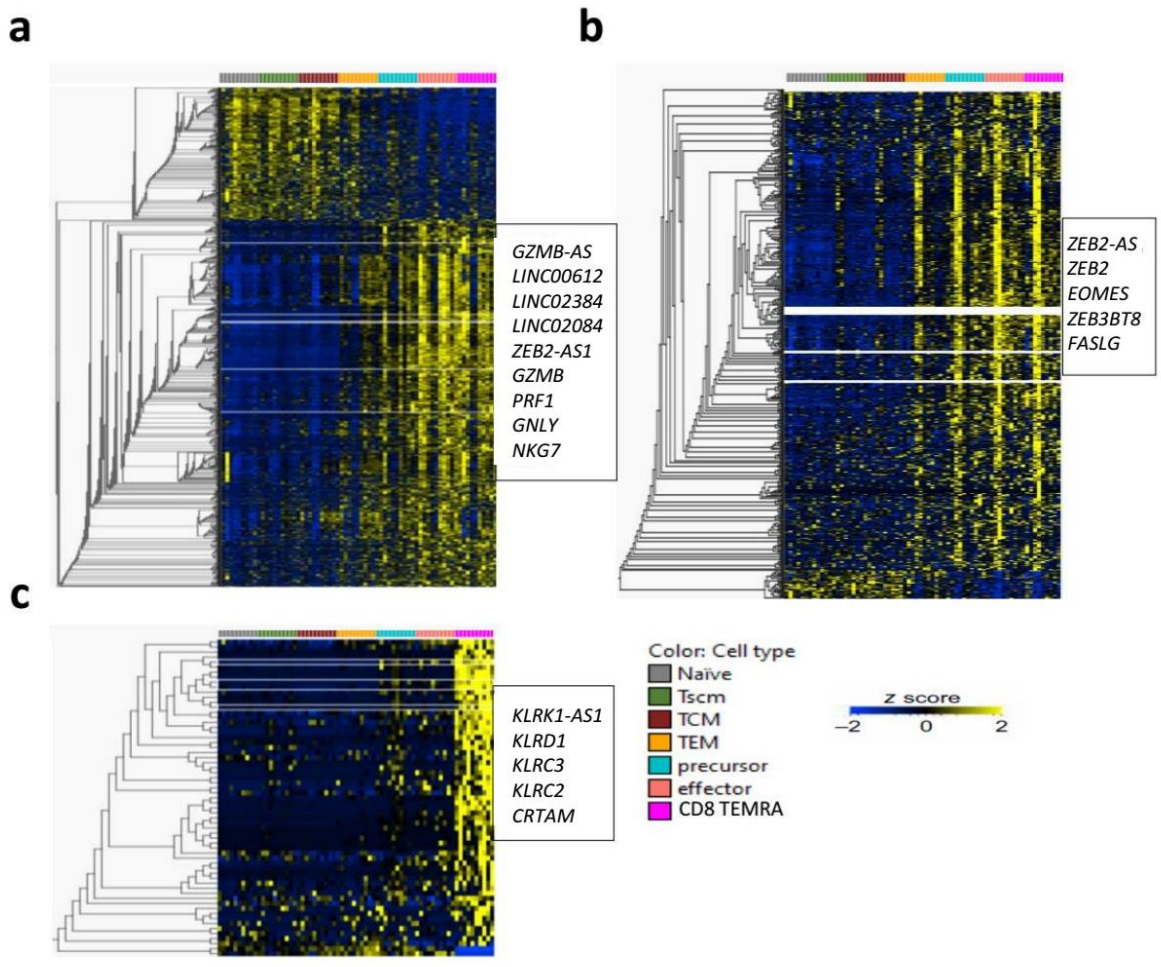


Figure 6: Heatmap depicting the modules of interest, yellow **(a)**, black **(b)** and darkorange2 **(c)** – specific to effector memory. lncRNAs coming in each module along with some effector-memory specific genes from the modules are highlighted. The genes are ordered according to hierarchical clustering and each row corresponds to a gene coloured by the normalized z-score. Samples (cell type) are arranged based on their effector profile.

lncRNA	Cis-gene	spearman correlation (r)
KLRK1-AS1	KLRK1	0.83
LINC00612	KLRG1	0.93
LINC02384	IFNG	0.86
LINC02084	EOMES	0.82
ZEB2-AS1	ZEB2	0.87
ZEB2-AS	ZEB2	0.85

Table 3. List of lncRNA-mRNA pairs selected for functional characterization. The corresponding correlation coefficient for each lncRNA-mRNA pair is given. All the correlation are statistically significant (P (two tailed) < 0.0001).

Additional details of the lncRNA-mRNA pairs selected for functional characterization

KLRK1-AS1 and KLRK1: KLRK1-AS1 is a lncRNA (4 splice variants) which is located on chromosome 12 and antisense to KLRK1. KLRK1 is an activating cell surface receptor predominantly expressed on cytotoxic immune cells like NK cells and CD8+ T cells, where their ligands are presented by stressed cells (López-Larrea et al., 2008). They have also been characterized as a co-stimulatory molecule of CD8+T cells (Groh et al., 2001). Recent studies have shown their involvement in CD4+ T cell differentiation in disease conditions (Zhou et al., 2023).

LINC02084 & EOMES: LINC02084 is a lncRNA which is located on chromosome 3 and 9.5 Kb away from the TSS of EOMES. EOMES, like T-bet (the Th1'master regulator'), belongs to the T-box transcription factor family (Papaioannou and Silver, 2003). While it was initially identified for its involvement in embryonic development, EOMES has now been shown to perform essential roles in the immune system, particularly in T cells. Studies on CD8+ T cells have

established their involvement in cytotoxic activity (Pearce et al., 2003), and recently, their involvement in CD4+ T cells has also been identified (Qui et al., 2011).

LINC02384 & IFN-γ: LINC02384 is a lncRNA (2 splice variants) which is located on chromosome 12 and 224 Kb away from the TSS of IFN-γ. IFN-γ is a cytokine secreted mainly by natural killer (NK) cells and activated T cells. It has the ability to stimulate the innate immune system, improve antigen presentation, and mediate antiviral and antibacterial immunity (Tau and Rothman, 1999). IFN-γ plays a crucial role in case of CD4+ T cells that facilitates the differentiation of naive CD4+ cells into effector Th1 T cells (Bradley et al., 1996; Serroukh et al., 2018).

LINC00612 & KLRG1: LINC00612 is a lncRNA (2 splice variants) which is located on chromosome 12 and 54 Kb away from KLRG1. KLRG1 is a co-inhibitory receptor found on natural killer cells and human T cells that have encountered antigens, which is a member of the lectin-like inhibitory killer cell receptor family (Ibegbu et al., 2005;). KLRG1 is conventionally used as a marker of senescence and is highly expressed by terminal effector T cells (Henson & Akbar, 2009). Apart from being a marker of senescence, they have been shown to be involved in additional signaling pathways in terminally differentiated T cells (Li et al., 2016; Hu ZhiDong et al., 2018)

ZEB2-AS1, ZEB2-AS and ZEB2: ZEB2-AS1 (4 splice variants) and ZEB2-AS (a novel transcript, antisense to ZEB2) is a lncRNA which is located on chromosome 2 and antisense to ZEB2. ZEB2 is a member of the zinc-finger E homeobox-binding (ZEB) protein family of transcription factors. They are

most recognized for suppressing epithelial genes to induce the epithelial-to-mesenchymal transition (EMT) (Vandewalle et al., 2005). Transcriptome analysis of CD8+ T cells has revealed that ZEB2 is highly expressed in effector cells (Gattinoni et al., 2011). And several studies have shown that ZEB2 probably functions in a broader transcriptional network with T-bet and ID2, controlling genes required for the development of the effector-memory T cell state while limiting the capacity of long-term memory cells (Omilusik et al 2015; Dominguez et al., 2015).

Regulation of lncRNA-mRNA pairs at the chromatin level

Nucleosome-free (de-repressed) region of the chromatin around a gene body might be an indication towards the transcriptional state of the gene - active, has been active recently, or going-to-be active. Simultaneously analyzing the open chromatin region along with the transcriptome will provide additional levels of information, such as identifying what is only poised but not expressed at the transcript level in a subset-specific manner. To obtain additional level of information about the lncRNA-mRNA pairs selected, tracks generated in the UCSC genome browser were used (using data from bulk ATAC-seq (openness of chromatin) & RNA seq of CD4+ T cell memory compartments) (Kar et al. unpublished) (fig 7).

LINC02084 and EOMES: Expression of this lncRNA-mRNA pair is effector memory-specific, as discussed previously. This is further visually confirmed based on the RNA tracks in the UCSC genome browser plot (bottom half). Interestingly, the

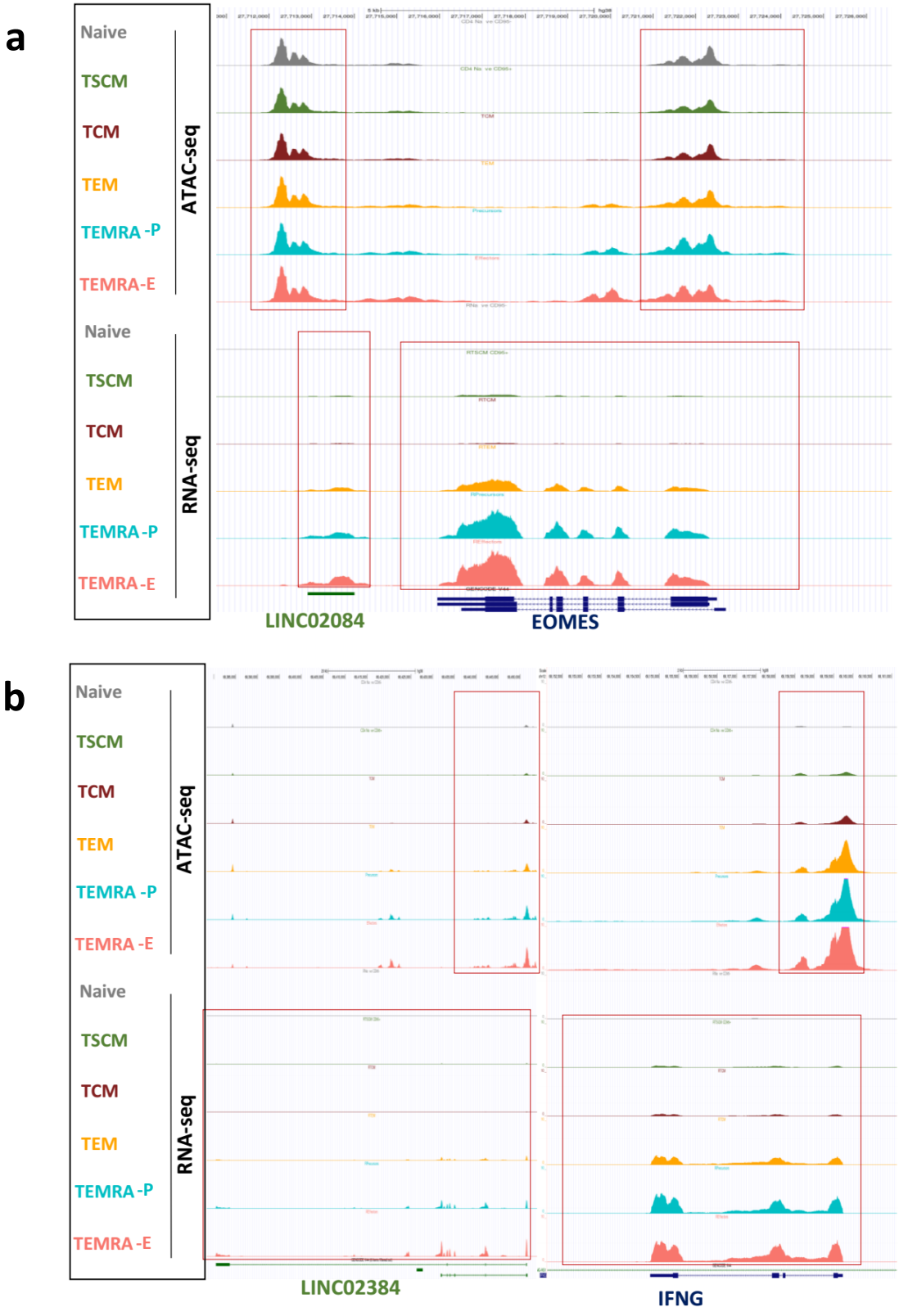
open chromatin analysis shows the TSS to be open across all the memory compartments irrespective of their effector status for both EOMES and LINC02084 (indicated by high-intensity peaks at the start site of EOMES and LINC02084). This hints toward a regulation by an element further upstream of this pair. Furthermore, though the intensity at TSS is not variable across the memory compartments, there is more openness at the exonic region of EOMES in the effector memory compartments. This could hint at the propensity of expression of EOMES upon regulation by its lncRNA counterpart (fig 7a).

LINC02384 & IFN-γ: This is also an effector pair discussed above, with further evidence from the RNA tracks of the UCSC genome browser. Openness around TSS gradually increases along with increasing effector-ness of memory compartments (Naive < TSCM < TCM < TEM < TEMRA-P < TEMRA-E). Further, as mRNA expression is observed in the effector memory compartments, the corresponding exonic region is open. This observation suggests the regulatory potential of LINC02384 with respect to IFN-γ expression in an effector-memory-specific scenario (fig 7b).

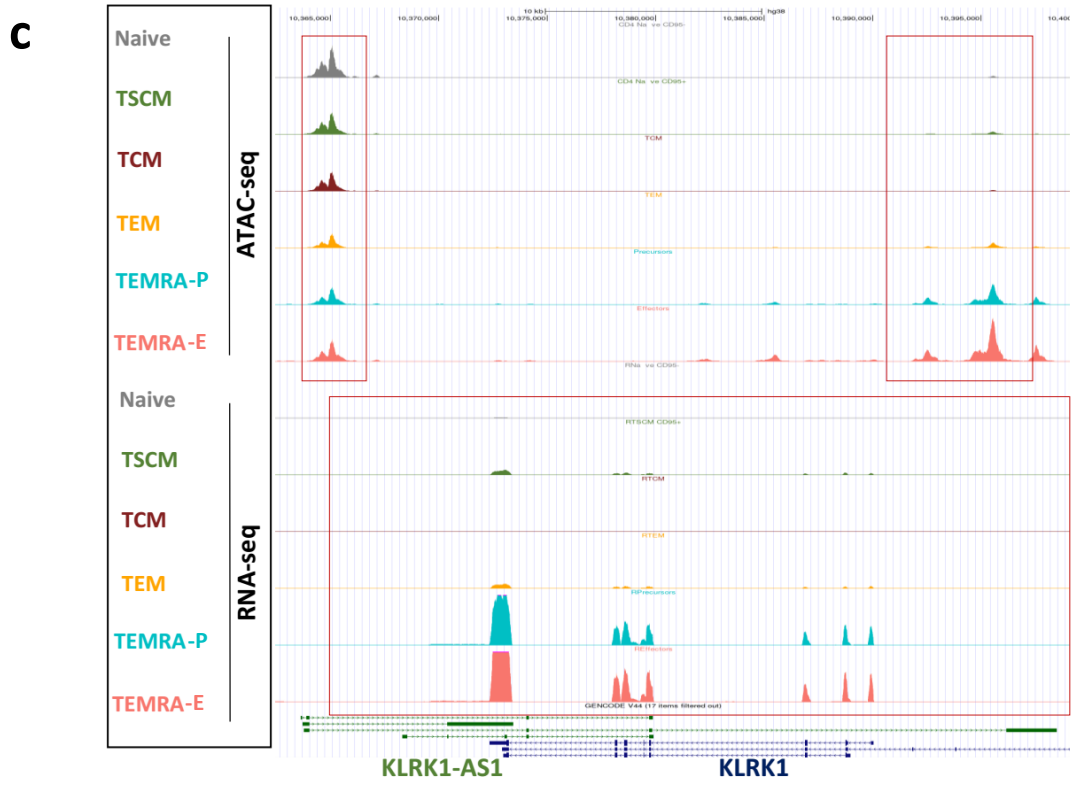
KLRK1-AS1 and KLRK1: Expression of this lncRNA-mRNA pair is TEMRA specific as discussed previously. Though the expression of KLRK1-AS1 was observed in effector-memory, the openness near the TSS is the opposite, having the highest intensity peak in the naive compartment. This suggests a negative feedback loop for the expression of the lncRNA where its expression negatively regulates the openness at TSS. Conversely, for KLRK1, the TSS and exonic region are open only in the TEMRA compartment, suggesting a plausible regulation that KLRK1 expression may be tightly controlled by KLRK1-AS1 (fig 7c).

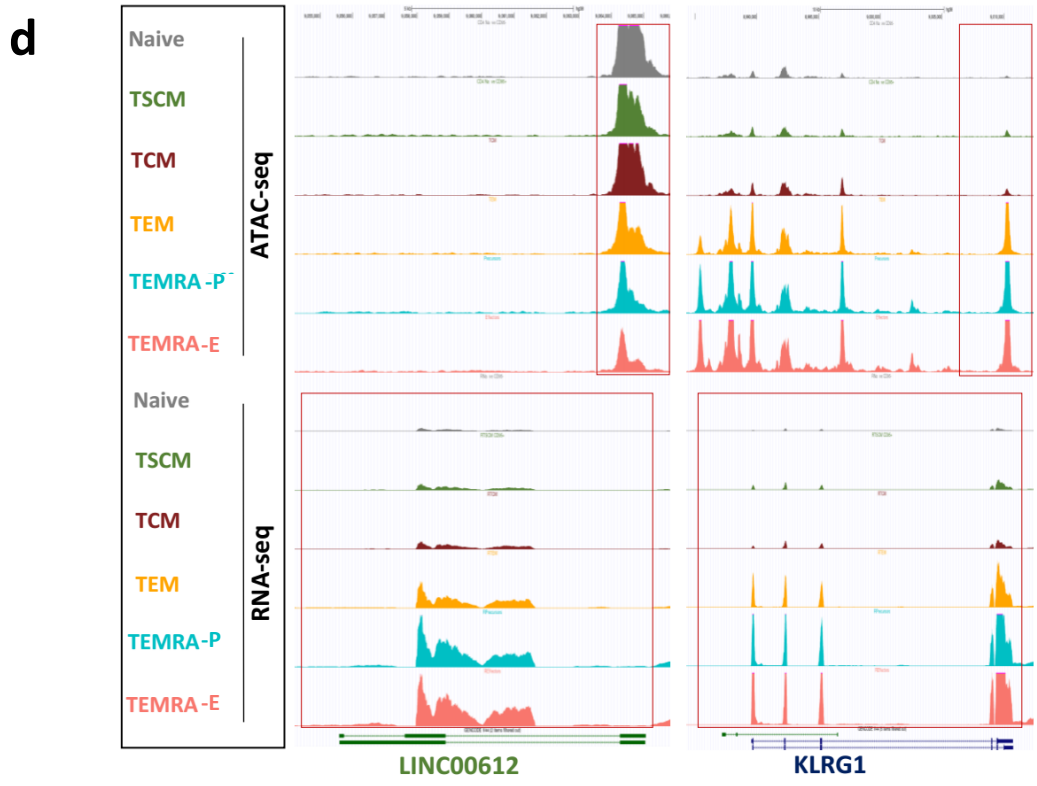
LINC00612 and KLRG1: Similar to the previous pair (KLRK1-AS1 and KLRK1), the dynamicity in the gene body openness suggests the limited expression of KLRG1 in the effector memory compartment (fig 7d).

ZEB2-AS1, ZEB2-AS, and ZEB2: the RNA expression level is high in effector memory and corroborates with openness of TSS. But the interesting thing to note here is the apparent mRNA expression in TSCM irrespective of lower levels of openness in chromatin (indicated by less intense peaks at TSS as compared to effector memory subsets). This warrants further significance testing of the peaks (fig 7e).



Legend in page 37





Legend in page 37

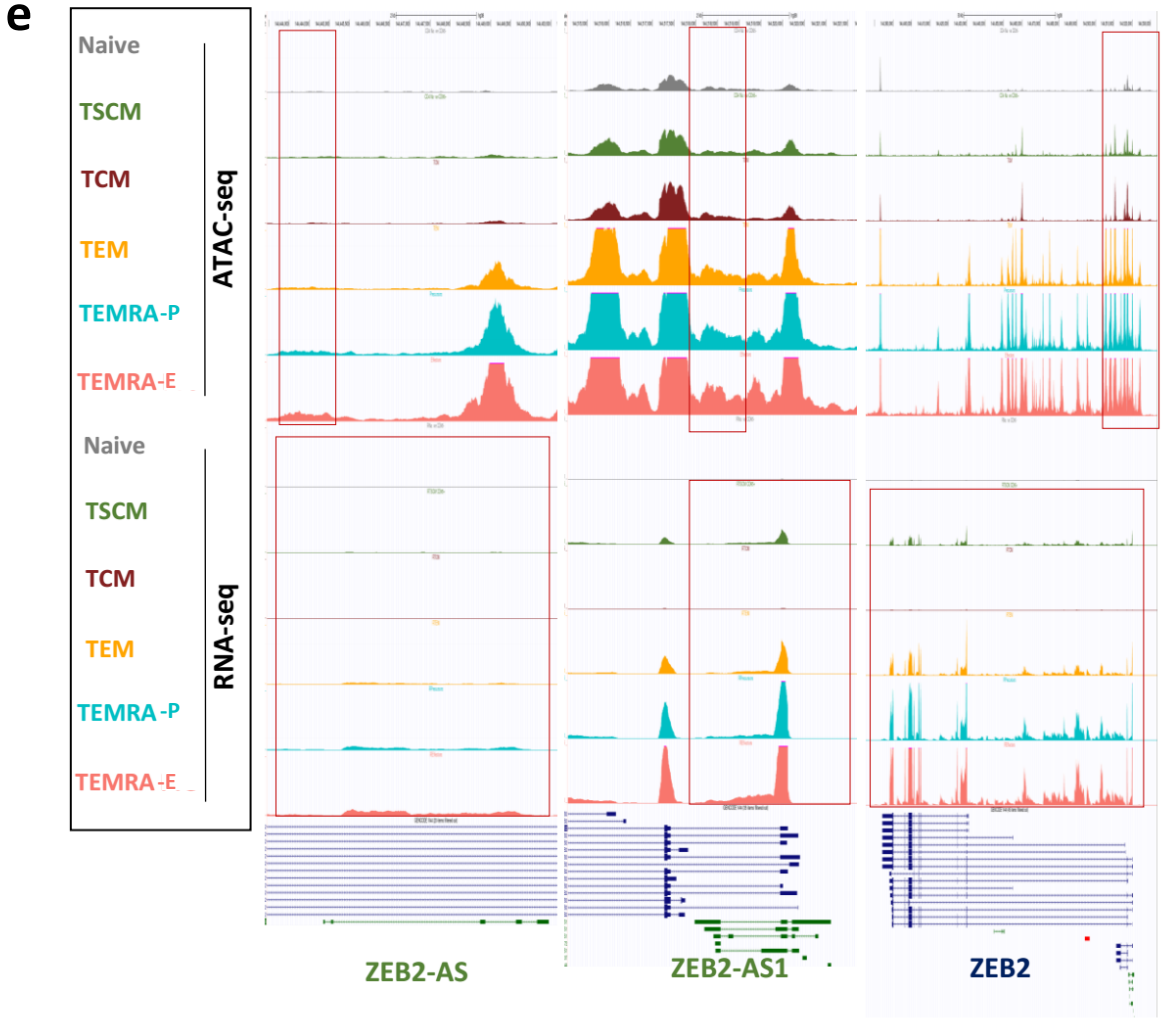


Figure 7. Tracks generated in the UCSC genome browser for lncRNA-mRNA pairs. Peaks corresponds to openness of chromatin (ATAC-seq) and RNA expression level (RNA-seq). Genome position of the lncRNA gene (green) and their corresponding protein-coding gene (dark blue) is given at the bottom. Transcript variants are coloured same. For ATAC-seq, openness corresponding to TSS is only labelled (red box). For RNA-seq, the entire exons are labelled and peak corresponds to respective exons represent the expression level. Higher the intensity of the peaks, higher the openness of chromatin and RNA expression level and vice versa.

Discussion

Large-scale changes occur through the course of memory development, at both transcriptomic as well as epigenetic levels. Previous studies have demonstrated that transcriptomic changes are not limited to protein-coding genes; non-coding genes were also differentially expressed (Ranzani et al., 2015; Hudson et al., 2019). lncRNAs, which represent a major class among the non-coding genes, have been shown to perform a wide array of functions and have been verified to play regulatory roles in innate and adaptive immune responses (Zhang et al., 2019; Petermann et al., 2019; Spurlock III et al., 2015). lncRNAs have also been shown to play regulatory roles in T memory development, but the studies were limited to conventional memory compartments (like TCM, TEM) and helper subsets (Th1, Th2, Th17, Treg) (Spurlock, et al., 2017; Ranzani et al., 2015). Here, I analysed the transcriptome of CD4+ T cell memory at a higher resolution by further dividing long-term memory compartments into TSCM and TCM, and effector-memory compartments into TEM, TEMRA-P, and TEMRA-E to extend the understanding of lncRNAs in CD4+ T cell memory compartments. Principal component analysis (PCA) on the transcriptome revealed that lncRNAs were differentially expressed across memory compartments like protein-coding genes, and they clustered the memory compartments into long-term memory and effector memory along the PC1. Even though both protein-coding and lncRNAs clustered the same way, The PC1 (32%) of protein-coding genes was high compared to the PC1 (20%) of lncRNAs. This could be attributed to the fact that lncRNAs are lowly expressed molecules as compared to mRNAs or the distinct expression of lncRNAs across memory is less pronounced as compared to mRNAs.

This observation of clustering the memory compartments led to further exploring the dataset to identify the strong lncRNA candidates associated with this. Considering the nature of the dataset I used, having representation from long-term memory and effector memory compartments, this study helped to capture strong lncRNA candidates that seem to be associated with longevity and effector profile of CD4 T cells. Therefore, the expression pattern of these memory specific lncRNAs might be useful in defining CD4+T cell memory compartments into long-term and effector memory compartments. Further studies need to be performed to verify their regulatory roles and to make a general statement that they are not limited to CD4+ T cells.

This study also opened up the scope for identifying potential cis-regulatory lncRNAs and protein-coding genes, where lncRNAs are known to have regulatory roles on their cis protein-coding genes (Spurlock III et al., 2015; Zhang et al., 2019). With the differential expression analysis and WGCNA, I identified effector-memory specific lncRNAs and paired some lncRNAs with their cis protein-coding genes. Interestingly, the lncRNAs selected are also expressed by other well-characterized effector immune cells (NK cells and effector CD8+ T cells), further supporting their role associated with the effector profile. Therefore, this study was able to capture effector-specific lncRNAs in the CD4+ T cell context, which is a less explored area. However, functional characterization is needed to verify their regulatory role.

lncRNAs are also known for their regulatory roles in various cellular processes and in disease conditions (Patil et al., 2014; Li et al., 2021). This hints at the notion that tight regulation of the lncRNAs is critical for normal physiological functioning. To explore to what degree this notion is applicable to the lncRNAs selected (for functional characterization), regulation at the

open chromatin level was explored. The analysis based on tracks generated in the UCSC genome browser showed that lncRNAs are tightly regulated at the chromatin level, and this might be responsible for the tight regulation of their corresponding protein-coding genes. For example, KLRK1-AS1-KLRK1 and LINC00612-KLRG1 showed a similar pattern where the lncRNA expression itself is tightly regulated irrespective of their TSS openness, resulting in their effector-specific expression pattern. However, the mentioned observation of regulation at the chromatin level is purely based on the tracks generated in the UCSC genome browser using ATAC-seq data, it demands further analysis of the ATAC-seq data in depth to make any conclusive statements.

Overall, this study contributed to adding an additional layer of information to the existing knowledge about the role of lncRNAs in the context of CD4+ T cell differentiation into different memory compartments. Importantly, this study led to the identification of strong candidate lncRNAs responsible for longevity and effector profile of CD4+ memory T cells. Additionally, this study identified potential cis-regulatory lncRNAs and protein-coding genes specific to effector-specific memory compartments which represent excellent targets for future functional studies.

References

- Araki, Y., Wang, Z., Zang, C., Wood, W. H., Schones, D., Cui, K., ... & Weng, N. P. (2009). Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. *Immunity*, *30*(6), 912-925.
- Bradley, L. M., Dalton, D. K., & Croft, M. (1996). A direct role for IFN-gamma in regulation of Th1 cell development. *Journal of immunology (Baltimore, Md.: 1950)*, *157*(4), 1350-1358.
- Chang, K. C., Diermeier, S. D., Yu, A. T., Brine, L. D., Russo, S., Bhatia, S., ... & Spector, D. L. (2020). MaTAR25 lncRNA regulates the Tensin1 gene to impact breast cancer progression. *Nature communications*, *11*(1), 6438.
- Chen, Y., Zander, R., & Cui, W. (2018). Transcriptional and epigenetic regulation of effector and memory CD8 T cell differentiation. *Frontiers in immunology*, *9*, 422590.
- Cheroutre, H., & Husain, M. M. (2013, November). CD4 CTL: living up to the challenge. In *Seminars in immunology* (Vol. 25, No. 4, pp. 273-281). Academic Press.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., ... & Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*, *22*(9), 1775-1789.
- Dominguez, C. X., Amezquita, R. A., Guan, T., Marshall, H. D., Joshi, N. S., Kleinstein, S. H., & Kaech, S. M. (2015). The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *Journal of Experimental Medicine*, *212*(12), 2041-2056.
- Eshima, K., Misawa, K., Ohashi, C., & Iwabuchi, K. (2018). Role of T-bet, the master regulator of Th1 cells, in the cytotoxicity of murine CD4+ T cells. *Microbiology and immunology*, *62*(5), 348-356.
- Gattinoni, L., Lugli, E., Ji, Y., Pos, Z., Paulos, C. M., Quigley, M. F., ... & Restifo, N. P. (2011). A human memory T cell subset with stem cell–like properties. *Nature medicine*, *17*(10), 1290-1297.
- Gloss, B. S., & Dinger, M. E. (2016). The specificity of long noncoding RNA expression. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1859*(1), 16-22.
- Groh, V., Rhinehart, R., Randolph-Habecker, J., Topp, M. S., Riddell, S. R., & Spies, T. (2001). Costimulation of CD8αβ T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nature immunology*, *2*(3), 255-260.
- Henson, S. M., & Akbar, A. N. (2009). KLRG1—more than a marker for T cell senescence. *Age*, *31*, 285-291.
- Hu ZhiDong, H. Z., Zhao HuiMin, Z. H., Li ChunLing, L. C., Liu XuHui, L. X., Barkan, D., Lowrie, D. B., ... & Fan XiaoYong, F. X. (2018). The role of KLRG1 in human CD4+ T-cell immunity against tuberculosis.
- Hudson, W. H., Prokhnevska, N., Gensheimer, J., Akondy, R., McGuire, D. J., Ahmed, R., & Kissick, H. T. (2019). Expression of novel long noncoding RNAs defines virus-specific effector and memory CD8+ T cells. *Nature communications*, *10*(1), 196.
- Ibegbu, C. C., Xu, Y. X., Harris, W., Maggio, D., Miller, J. D., & Kourtis, A. P. (2005). Expression of killer cell lectin-like receptor G1 on antigen-specific human CD8+ T lymphocytes during active, latent, and resolved infection and its relation with CD57. *The Journal of Immunology*, *174*(10), 6088-6094.
- ieberman, J. (2003). The ABCs of granule-mediated cytotoxicity: new weapons in the arsenal. *Nature Reviews Immunology*, *3*(5), 361-370.
- Juno, J. A., Van Bockel, D., Kent, S. J., Kelleher, A. D., Zaunders, J. J., & Munier, C. (2017). Cytotoxic CD4 T cells—Friend or foe during viral infection?. *Frontiers in immunology*, *8*, 242741.
- Kaech, S. M., & Cui, W. (2012). Transcriptional control of effector and memory CD8+ T cell differentiation. *Nature Reviews Immunology*, *12*(11), 749-761.

- Kaech, S. M., Hemby, S., Kersh, E., & Ahmed, R. (2002). Molecular and functional profiling of memory CD8 T cell differentiation. *Cell*, *111*(6), 837-851.
- Li, G., Deng, L., Huang, N., & Sun, F. (2021). The biological roles of lncRNAs and future prospects in clinical application. *Diseases*, *9*(1), 8.
- Li, L., Wan, S., Tao, K., Wang, G., & Zhao, E. (2016). KLRG1 restricts memory T cell antitumor immunity. *Oncotarget*, *7*(38), 61670.
- López-Larrea, C., Suárez-Alvarez, B., López-Soto, A., López-Vázquez, A., & Gonzalez, S. (2008). The NKG2D receptor: sensing stressed cells. *Trends in molecular medicine*, *14*(4), 179-189.
- Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M., & Lugli, E. (2013). The who's who of T-cell differentiation: human memory T-cell subsets. *European journal of immunology*, *43*(11), 2797-2809.
- Mattick, J. S., Amaral, P. P., Carninci, P., Carpenter, S., Chang, H. Y., Chen, L. L., ... & Wu, M. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nature reviews Molecular cell biology*, *24*(6), 430-447.
- Omilusik, K. D., Best, J. A., Yu, B., Goossens, S., Weidemann, A., Nguyen, J. V., ... & Goldrath, A. W. (2015). Transcriptional repressor ZEB2 promotes terminal differentiation of CD8+ effector and memory T cell populations during infection. *Journal of Experimental Medicine*, *212*(12), 2027-2039.
- Papaioannou, V. E., & Silver, L. M. (1998). The T-box gene family. *Bioessays*, *20*(1), 9-19.
- Patil, V. S., Madrigal, A., Schmiedel, B. J., Clarke, J., O'Rourke, P., de Silva, A. D., ... & Vijayanand, P. (2018). Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Science immunology*, *3*(19), eaan8664.
- Patil, V. S., Zhou, R., & Rana, T. M. (2014). Gene regulation by non-coding RNAs. *Critical reviews in biochemistry and molecular biology*, *49*(1), 16-32.
- Pearce, E. L., Mullen, A. C., Martins, G. A., Krawczyk, C. M., Hutchins, A. S., Zediak, V. P., ... & Reiner, S. L. (2003). Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science*, *302*(5647), 1041-1043.
- Petermann, F., Pękowska, A., Johnson, C. A., Jankovic, D., Shih, H. Y., Jiang, K., ... & O'Shea, J. J. (2019). The magnitude of IFN-γ responses is fine-tuned by DNA architecture and the non-coding transcript of Ifng-as1. *Molecular cell*, *75*(6), 1229-1242.
- Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols*, *9*(1), 171-181.
- Qui, H. Z., Hagymasi, A. T., Bandyopadhyay, S., St Rose, M. C., Ramanarasimhaiah, R., Ménoret, A., ... & Adler, A. J. (2011). CD134 plus CD137 dual costimulation induces Eomesodermin in CD4 T cells to program cytotoxic Th1 differentiation. *The Journal of Immunology*, *187*(7), 3555-3564.
- Ranzani, V., Rossetti, G., Panzeri, I., Arrigoni, A., Bonnal, R. J., Curti, S., ... & Pagani, M. (2015). The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nature immunology*, *16*(3), 318-325.
- Salehi, S., Taheri, M. N., Azarpira, N., Zare, A., & Behzad-Behbahani, A. (2017). State of the art technologies to explore long non-coding RNAs in cancer. *Journal of cellular and molecular medicine*, *21*(12), 3120-3140.
- Sallusto, F., Geginat, J., & Lanzavecchia, A. (2004). Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.*, *22*, 745-763.
- Sallusto, F., Lenig, D., Förster, R., Lipp, M., & Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*, *401*(6754), 708-712.
- Serroukh, Y., Gu-Trantien, C., Hooshiar Kashani, B., Defrance, M., Vu Manh, T. P., Azouz, A., ... & Marchant, A. (2018). The transcription factors Runx3 and ThPOK cross-regulate acquisition of cytotoxic function by human Th1 lymphocytes. *Elife*, *7*, e30496.
- Spurlock III, C. F., Tossberg, J. T., Guo, Y., Collier, S. P., Crooke III, P. S., & Aune, T. M. (2015). Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nature communications*, *6*(1), 6932.

- Spurlock, C. F., Shaginurova, G., Tossberg, J. T., Hester, J. D., Chapman, N., Guo, Y., ... & Aune, T. M. (2017). Profiles of long noncoding RNAs in human naive and memory T cells. *The Journal of Immunology*, *199*(2), 547-558.
- Tau, G., & Rothman, P. (1999). Biologic functions of the IFN-γ receptors. *Allergy*, *54*(12), 1233.
- Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*, *154*(1), 26-46.
- Vandewalle, C., Comijn, J., De Craene, B., Vermassen, P., Bruyneel, E., Andersen, H., ... & Berx, G. (2005). SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell–cell junctions. *Nucleic acids research*, *33*(20), 6566-6578.
- Watson, A. M., Lam, L. M., Klimstra, W. B., & Ryman, K. D. (2016). The 17D-204 vaccine strain-induced protection against virulent yellow fever virus is mediated by humoral immunity and CD4+ but not CD8+ T cells. *PLoS pathogens*, *12*(7), e1005786.
- Weiskopf, D., Bangs, D. J., Sidney, J., Kolla, R. V., De Silva, A. D., de Silva, A. M., ... & Sette, A. (2015). Dengue virus infection elicits highly polarized CX3CR1+ cytotoxic CD4+ T cells associated with protective immunity. *Proceedings of the National Academy of Sciences*, *112*(31), E4256-E4263.
- Youngblood, B., Hale, J. S., & Ahmed, R. (2013). T-cell memory differentiation: insights from transcriptional signatures and epigenetics. *Immunology*, *139*(3), 277-284.
- Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*, *4*(1).
- Zhang, J., Cheng, Y., Cui, W., Li, M., Li, B., & Guo, L. (2014). MicroRNA-155 modulates Th1 and Th17 cell differentiation and is associated with multiple sclerosis and experimental autoimmune encephalomyelitis. *Journal of neuroimmunology*, *266*(1-2), 56-63.
- Zhang, Q., Chao, T. C., Patil, V. S., Qin, Y., Tiwari, S. K., Chiou, J., ... & Rana, T. M. (2019). The long noncoding RNA ROCKI regulates inflammatory gene expression. *The EMBO journal*, *38*(8), e100041.
- Zhou, J., Wang, J., Tao, L., Liu, M., Tang, X., & Zhu, X. (2023). NKG2D receptor regulates CD4+ T cell differentiation via interaction with dendritic cells in patients with juvenile idiopathic arthritis. *Clinical Immunology*, *256*, 109780.