Genomics Data 2 (2014) 396–398

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Whole transcriptome analysis for T cell receptor-affinity and IRF4-regulated clonal expansion of T cells

Wei Shi [a,b,*], Kevin Man [a,c], Gordon K. Smyth [a,d], Stephen L. Nutt [a,c], Axel Kallies [a,c]

[a] The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia
[b] The Department of Computing and Information Systems, University of Melbourne, Parkville, Australia
[c] The Department of Medical Biology, University of Melbourne, Parkville, Australia
[d] The Department of Mathematics and Statistics, University of Melbourne, Parkville, Australia

ARTICLE INFO

Article history:
Received 21 October 2014
Accepted 28 October 2014
Available online 1 November 2014

Keywords:
IRF4
RNA-sequencing analysis
TCR-affinity
T cells
Clonal expansion

ABSTRACT

Clonal population expansion of T cells during an immune response is dependent on the affinity of the T cell receptor (TCR) for its antigen [1]. However, there is little understanding of how this process is controlled transcriptionally. We found that the transcription factor IRF4 was induced in a manner dependent on TCR-affinity and was critical for the clonal expansion and maintenance of effector function of antigen-specific CD8$^+$ T cells. We performed a genome-wide expression profiling experiment using RNA sequencing technology (RNA-seq) to interrogate global expression changes when IRF4 was deleted in CD8$^+$ T cells activated with either a low or high affinity peptide ligand. This allowed us not only to determine IRF4-dependent transcriptional changes but also to identify transcripts dependent on TCR-affinity [2]. Here we describe in detail the analyses of the RNA-seq data, including quality control, read mapping, quantification, normalization and assessment of differential gene expression. The RNA-seq data can be accessed from Gene Expression Omnibus database (accession number GSE49929).

Specifications	
Organism/cell line/tissue	*Mus musculus*/spleen and lymph node
Strain	C57BL/6
Sequencer or array type	Illumina HiSeq 2000 sequencer
Data format	FASTQ
Experimental factors	Wild-type or IRF4$^{-/-}$ CD8$^+$ OT-1 T cells, activated with either a low or high affinity peptide ligand
Experimental features	RNA-seq data
Consent	Mice were maintained and used in accordance with the guidelines of the Walter and Eliza Hall Institute Animal Ethics Committee.
Sample source location	Melbourne, Australia

Direct link to deposited data

Deposited data can be accessed via: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49929.

Experimental design, materials and methods

Sample preparation

Irf4$^{-/-}$ mice have been described in [3] and were maintained on a C57BL/6 (Ly5.2$^+$) background. They were crossed to OT-I mice, which carry a MHCI-restricted TCR-transgene resulting in the expression of an ovalbumin (OVA) peptide specific TCR [4]. Naive CD8$^+$ T cells were isolated from the spleens and lymph nodes of OT-I mice on either a wild-type or *Irf4*$^{-/-}$ background and were activated for 72 h in vitro with OVA peptide N4 (SIINFEKL, high affinity) or V4 (SIIVFEKL, low affinity) [1] (1 ug/ml) in the presence of recombinant human IL-2 (100 U/ml; R&D Systems).

RNA sequencing

RNA was purified with an RNAeasy Plus Mini Kit according to the manufacturer's protocol (Qiagen). The DNA fragments were ligated to Illumina adaptors with blunt ends and were amplified, then were sequenced with an Illumina HiSeq 2000 sequencer. Each sample had two or three biological replicates. Paired-end 90 bp reads were generated from sequencing.

[*] Corresponding author at: The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia.
E-mail address: shi@wehi.edu.au (W. Shi).

http://dx.doi.org/10.1016/j.gdata.2014.10.019

Sequencing quality

Fig. 1 shows the distribution of base-calling Phred scores at each base location in all the reads included in one of the libraries. Although nucleotides located at the ends of reads were found to have a lower sequencing quality than those in the middle of reads, the overall sequencing quality is high since the majority of read bases have a Phred score greater than 30 (ie. probability of incorrect base calling is less than 0.001). Other libraries included in this study were found to have a sequencing quality similar to that shown in Fig. 1.

Read mapping and summarization

Sequence reads were mapped to mouse reference genome *mm9* using the Subread aligner [5], which is capable of mapping both exonic and exon-spanning reads. Mapped reads were summarized to NCBI RefSeq genes using the featureCounts program [6]. Raw read counts were generated for each gene in each library after summarization.

Gene filtering and normalization

Genes were removed from the analysis if they failed to achieve a FPKM (fragments per kilobases per million mapped reads) value of 0.5 or greater in at least one library. Counts were converted to log2 counts per million (CPM), quantile normalized and precision weighted using voom [7]. Fig. 2 shows the relationship between mean expression values of genes and their expression variations. Expression variations of genes were estimated from the biological replicates. Fig. 3 shows the clustering of samples after normalization. Distinct cell types were clearly separated and sample replicates were clustered together.

Differential expression analysis

Linear models were fitted to genes using the Bioconductor R package limma [8]. Precision weights for genes that were estimated by voom were used in the linear modeling process. Empirical Bayes moderated *t*-statistics were used to assess differential expression [9]. A false discovery rate of 5% and a fold change cutoff of 2 fold were applied for calling differentially expressed (DE) genes. Also, DE genes must have a FPKM value of 8 or greater in one or both of two samples being compared. DE genes found in comparing *Irf4*$^{-/-}$ with wild type in high-affinity CD8$^+$ T cells are highlighted in Fig. 4, in which genome-wide expression changes between the two samples are shown.

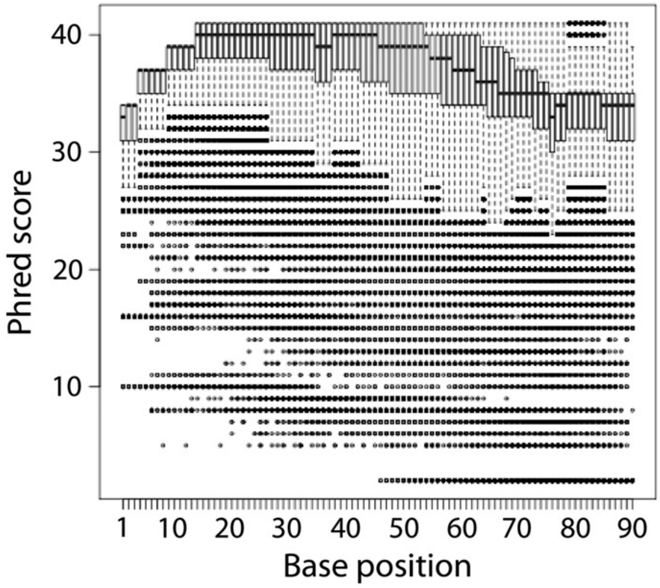

Fig. 1. Distribution of base-calling Phred scores at each base location in all the reads included in one of the libraries. The horizontal axis gives the position of each nucleotide in the read and the vertical axis shows a box plot of Phred scores of called nucleotides at each read position. For each base position, the box shows the 25%, 50% and 75% quantiles of the Phred scores. Scores more than 1.5 interquartile ranges from the median for that position are plotted as individual points. Phred scores of read bases were retrieved from the FASTQ input file using the *qualityScores* function in Bioconductor R package *Rsubread*.

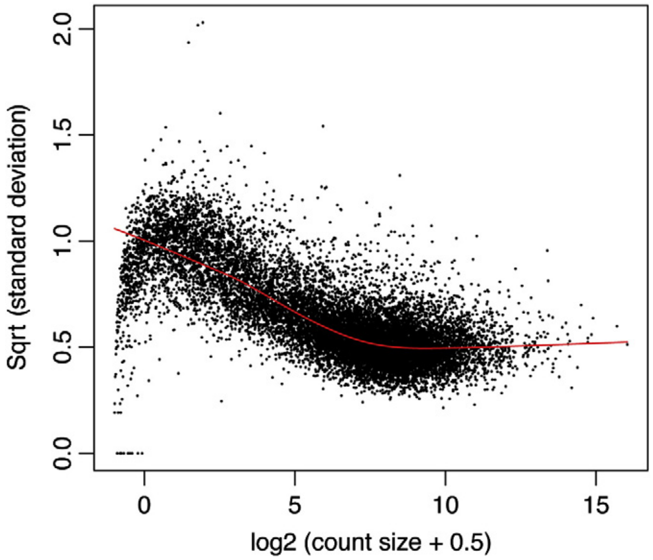

Fig. 2. Mean–variance relationship estimated from the sequence data by voom. The horizontal axis gives the mean log2-CPM values of genes and the vertical axis gives the square root of standard deviation of log2-CPM expression values of genes that is estimated from the biological replicates of samples.

Discussion

Here we provided a detailed description to the analyses we carried out for the RNA-seq data generated in the original study of TCR-affinity and IRF4-mediated transcriptional changes in CD8 T cells [2]. Raw sequence read data have been made publicly available and

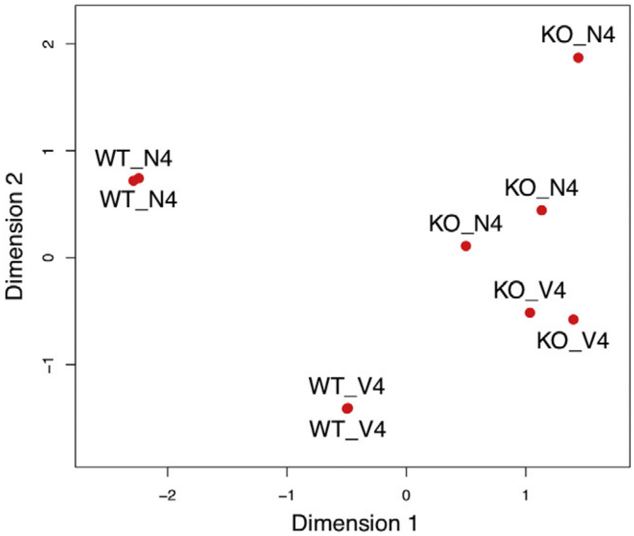

Fig. 3. Unsupervised clustering of the samples by multi-dimensional scaling. 'WT' and 'KO' denote wild-type and *Irf4*$^{-/-}$ OT-1 T cells, respectively. 'N4' and 'V4' denote stimulation with high affinity peptide and stimulation with low affinity peptide, respectively. Distances on the plot represent average absolute $\log_2$ fold change for the leading 500 genes that distinguish each pair of samples. This figure was generated using the *plotMDS* function in Bioconductor R package *limma*.

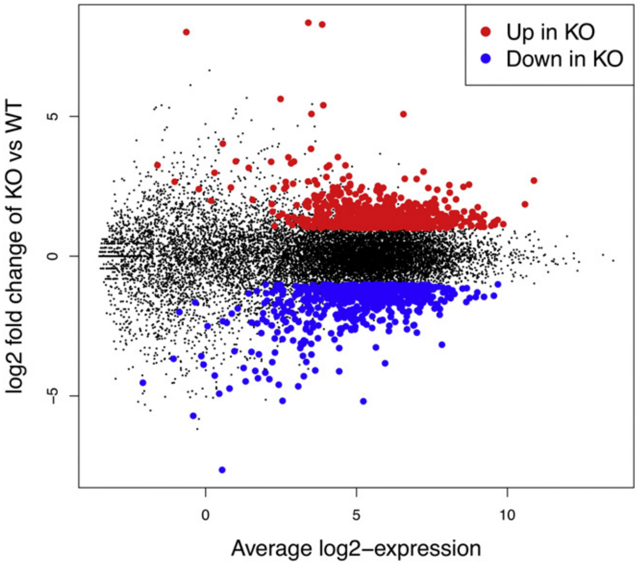

Fig. 4. Genome-wide expression changes between *Irf4*$^{-/-}$ and wild type in high-affinity CD8^{+} OT-1 T cells. Significantly up-regulated and down-regulated genes are highlighted. This figure was generated using the *plotMA* function in *limma*.

software programs used in this analysis can also be freely downloaded from Bioconductor [10] or SourceForge (http://subread.sourceforge.net). These should enable the RNA-seq analysis results presented in the original study to be readily reproduced. We also want to note that the pipeline used in this data analysis has been found to be one of the best-performing pipelines for RNA-seq analysis by the SEQC/MAQC III Consortium in their recent efforts to benchmark RNA-seq technologies [11].

Acknowledgements

This work was supported by grants and fellowships from the National Health and Medical Research Council of Australia (NHMRC) (GKS, SLN, AK, WS), the Sylvia and Charles Viertel Foundation (AK), the Australian Research Council (SLN, AK), and the WEHI Genomics Fund (AK). The funding includes NHMRC Project Grant 1023454 (GKS, WS), Project Grant 1032850 (AK) and Program grant 1054618 (GKS). This study was made possible through the Victorian State Government Operational Infrastructure Support and the Australian Government NHMRC Independent Research Institute Infrastructure Support scheme.

References

[1] D. Zehn, S.Y. Lee, M.J. Bevan, Complete but curtailed T-cell response to very low-affinity antigen. Nature 458 (2009) 211–214.

[2] K. Man, M. Miasari, W. Shi, A. Xin, D.C. Henstridge, et al., The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells. Nat. Immunol. 14 (2013) 1155–1165.

[3] H.W. Mittrucker, T. Matsuyama, A. Grossman, T.M. Kundig, J. Potter, et al., Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. Science 275 (1997) 540–543.

[4] K.A. Hogquist, S.C. Jameson, W.R. Heath, J.L. Howard, M.J. Bevan, et al., T cell receptor antagonist peptides induce positive selection. Cell 76 (1994) 17–27.

[5] Y. Liao, G.K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 41 (2013) e108.

[6] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30 (2014) 923–930.

[7] C.W. Law, Y. Chen, W. Shi, G.K. Smyth, Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15 (2014) R29.

[8] G.K. Smyth, Limma: linear models for microarray data. in: R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer, New York, 2005, pp. 397–420.

[9] G.K. Smyth, Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3 (2004) (Article3).

[10] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, et al., Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5 (2004) R80.

[11] Z. Su, P.P. Labaj, S. Li, J. Thierry-Mieg, D. Thierry-Mieg, et al., A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat. Biotechnol. 32 (2014) 903–914.