

Characterization of T Cell Reprogramming for Adoptive Therapy

Yiqun Chen*, Daniel Lee*, Alyssa Morrow*, James Kaminski¹, and Nir Yosef¹

*these authors contributed equally to this work

ABSTRACT

In adoptive therapy, T cells have been genetically modified to express Chimeric Antigen Receptors (CARs), used to target and kill cells presenting malignant patient-specific epitopes on the cell surface. While this method has been successful in treating common malignancies, we are left with the question of how to efficiently generate these CAR expressing T cells in large quantities. To produce a self-renewable source of these engineered cells, T cells have been reprogrammed to induced pluripotent stem cells (iPSC), modified to express a CAR, easily cloned and re-differentiated into T cells. In this paper, we explore the change in transcriptomic and epigenetic landscapes upon reprogramming of T lymphocytes to iPSC cells, hoping to gain insight into footprints that are left behind after the reprogramming process.

1 Introduction

Chimeric Antigen Receptor (CAR) expressing T cells have been used in adoptive therapy to target patient-specific malignant cells. One example in which adoption of CAR expressing T cells in tumor patients is seen in the case of acute myloid leukemia, in which patients have experienced 90% complete response rates¹. More specifically, stem memory T cells (Tsm) have had particularly strong anti-tumor effects, compared to T cells at other differentiation stages². Current methods of adoptive therapy with CAR expressing T cells, however, do not reap the benefits of Tsm anti-tumor effects, as ex vivo expanded T cells quickly differentiate to T effector cells, losing their phenotypic traits and subsequently driving differentiation towards cell death^{3,4}. Other methods for T cell expansion are not self-renewable, and thus it is difficult to generate a substantial amount of cells required for adoptive therapy.

However, a new method for ex vivo T cell generation reprograms T cells to iPSC cells, where the reprogrammed cells are modified with a CAR then expanded and re-differentiated to T cells⁵. T cells are reprogrammed to iPSC cells with the incorporation of Yamanaka factors and cultured with human T cells to induce pluripotency. The main benefit of this approach is self-renewal; iPSC cells are easily expanded ex vivo to be re-differentiated for further use. Other potential benefits of T cell reprogramming involve maintaining transcriptomic and epigenetic landscape of T cells. In this paper, we investigate this claim, measuring the change in gene expression and epigenetic landscape before and after reprogramming of naive and stem memory T cells to iPSC cells. We use both edgeR⁶ and DESeq2⁷ to measure differential expression from RNA-seq and differential accessibility from ATAC-seq among the reprogrammed cells. We conclude that the edgeR method is more sensitive to differential expression in T cell and pluripotent specific genes and therefore use this method for subsequent analyses. Lastly, we investigate differentiation of transcription factors that are known to regulate the process of T cell differentiation, and find that while drivers in T cell differentiation, such as Prdm1, are suppressed after reprogramming, other transcription factors that negatively regulate differentiation are removed after reprogramming.

2 DataSets

RNA-seq and ATAC-seq data were collected from 12 samples, from a total of four donors. The datasets collected can be summarized in Table. This dataset consists of two initial samples for both naive and stem memory cells, as well as four samples of both naive and stem memory reprogrammed to iPSC cells. Table 1 shows the comprehensive list of donors and cell types. Two donors for each of the cell categories was obtained for RNA- and ATAC-seq datasets. T cells had two samples for both naive and stem memory cells. iPSC reprogrammed cells had four samples for both naive and stem memory cells.

2.1 Reprogramming of T Lymphocytes to iPSC Cells

Five Sendai viruses were integrated with a separate Yamanaka factors, including MYC, Oct4, Sox2, KLF4 and GFP. For collection of reprogrammed T naive cells, CD8 Tn cells were exposed to these five Sendai viruses and incubated for 6 days. Similarly, 45,000 Tsm cells were incubated with the same Sendai viruses. Resulting cells for each sample were then evaluated using ATAC-seq and RNA sequencing.

ID	SampleName	CellType	Donor	Batch	ReplicateGroups
PT35_Tn_donor_3_input	R10_S16_R1_001	PT35_Tn_donor_3_input	Naive_T_Cell	Donor_3	BatchOne D3_Naive
PT35Tn_3_V5_iPSC	R11_S17_R1_001	PT35Tn_3_V5_iPSC	iPSC_Tn	Donor_3	BatchOne D3_iPSC_Tn
PT35Tn_3_V3_iPSC	R12_S18_R1_001	PT35Tn_3_V3_iPSC	iPSC_Tn	Donor_3	BatchOne D3_iPSC_Tn
PT32_Tsm_input_donor	R1_S7_R1_001	PT32_Tsm_input_donor	T_Stem_Memory Cell	Donor_1	BatchOne D1_TStemMem
PT32_Tsm_V1_iPSC	R2_S8_R1_001	PT32_Tsm_V1_iPSC	iPSC_Tsm	Donor_1	BatchOne D1_iPSC_Tsm
PT32_Tsm_V3_iPSC	R3_S9_R1_001	PT32_Tsm_V3_iPSC	iPSC_Tsm	Donor_1	BatchOne D1_iPSC_Tsm
PT33_Tsm_input_donor	R4_S10_R1_001	PT33_Tsm_input_donor	T_Stem_Memory Cell	Donor_4	BatchOne D4_TStemMem
PT33_Tsm_V2_iPSC	R5_S11_R1_001	PT33_Tsm_V2_iPSC	iPSC_Tsm	Donor_4	BatchOne D4_iPSC_Tsm
PT33_Tsm_V10_iPSC	R6_S12_R1_001	PT33_Tsm_V10_iPSC	iPSC_Tsm	Donor_4	BatchOne D4_iPSC_Tsm
PT35_Tn_donor_2_input	R7_S13_R1_001	PT35_Tn_donor_2_input	Naive_T_Cell	Donor_2	BatchOne D2_Naive
PT35_Tn_2_V4_iPSC	R8_S14_R1_001	PT35_Tn_2_V4_iPSC	iPSC_Tn	Donor_2	BatchOne D2_iPSC_Tn
PT35_Tn_2_V9_iPSC	R9_S15_R1_001	PT35_Tn_2_V9_iPSC	iPSC_Tn	Donor_2	BatchOne D2_iPSC_Tn

Figure 1. Table of sample groups. Four cell types were obtained from two donors per cell type.

3 Methods

3.1 edgeR

The following are the methods we implemented using the edgeR package in R.

3.1.1 Model

We fit a negative binomial regression model to four cell types across 12 samples using the `glmFit()` function. For a given sample i and region of interest j , the model is the following:

$$\log(E[Y_{i,j}|X]) = \beta_0 + \beta_1 \times Tsm + \beta_2 \times iPSCtN + \beta_3 \times iPSCtSmVar(Y_{i,j}|X) = E[Y_{i,j}|X] + \phi(E[Y_{i,j}|X])^2 \quad (1)$$

$$Var(Y_{i,j}|X) = E[Y_{i,j}|X] + \phi(E[Y_{i,j}|X])^2 \quad (2)$$

where ϕ is the dispersion parameter and Tsm , $iPSCtN$, and $iPSCtSm$ are indicator variables.

Using `estimateGLMTrendedDisp()`, we estimate the dispersion parameter, ϕ , for each gene or ATAC cut site with a trend that depends on the overall level of expression or accessibility of a region. This is done for a DGE dataset for general experimental designs by using Cox-Reid approximate conditional inference for a negative binomial generalized linear model for each gene with the unadjusted counts and design matrix provided.

For each gene, our design matrix will have 12 rows representing the 12 samples, and 4 columns, with the first column as the intercept, which represents our baseline cell category Tn, and the next three columns representing the three cell categories Tsm, iPSCtN, and iPSCtSm.

3.1.2 Hypothesis Test for Differential Expression

We use the `glmLRT()` function in edgeR to perform likelihood ratio tests and their corresponding p -values for each gene. P -values are adjusted to control the false discovery rate by the method of Benjamini and Hochberg. A gene is determined to be differentially expressed if the adjusted p -value for the estimated parameter value is smaller than 0.05. We performed the following three hypothesis tests for each gene:

$$\beta_1 = 0$$

$$\beta_2 = 0$$

$$\beta_1 = \beta_3$$

where $\beta*$ refers to the parameters defined in Equation 1.

3.2 DESeq2

The following are the methods we implemented using the edgeR package in R.

3.2.1 Filtering

DESeq2 automatically determines a threshold, filtering on mean normalized count, which maximizes the number of genes which will have an adjusted p -value less than a critical value of 0.05. This automatic independent filtering is performed by, and can be controlled by, the `results()` function.

3.2.2 Normalization

The function `counts()` has a parameter *normalized*. We set this parameter to *TRUE*. This will cause DESeq2 to perform normalization. The default normalization setting in DESeq2 is to divide each column of the count table by a normalization factor s_{ij} to account for the differences in sequencing depth between samples. To estimate this normalization factor, DESeq2 takes the median of the ratios of the observed counts. To estimate the median of the ratios of the observed counts, \hat{s}_j , DESeq2 performs Equation 3:

$$\hat{s}_j = \text{median}_i \frac{k_{ij}}{\left(\prod_{v=1}^m k_{iv}\right)^{1/m}} \quad (3)$$

where k_{ij} is the raw gene count in gene i sample j and m is the total number of samples. The denominator of this expression can be interpreted as a pseudo-reference sample obtained by taking the geometric mean across samples.

3.2.3 Model

Using the `DESeq()` function in DESeq2, we fit a negative binomial regression model. DESeq2 models the read counts K_{ij} with a negative binomial distribution with mean μ_{ij} and dispersion α_i , shown in Equation 4.

$$K_{ij} \sim NB(\mu_{ij}, \alpha_i) \quad (4)$$

The mean is taken as a quantity q_{ij} proportional to the concentration of cDNA fragments from the gene in the sample, scaled by a normalization factor s_{ij} . That is,

$$\mu_{ij} = s_{ij} q_{ij} \quad (5)$$

Variance of the counts is as follows:

$$Var K_{ij} = \mu_{ij} + \alpha_i \mu_{ij}^2 \quad (6)$$

Broadly speaking, DESeq2 uses the empirical Bayes shrinkage for dispersion estimation. Details on how this is done can be found in Love, Huber, Anders 2014⁸.

The link function for the GLM is \log_2 function. Hence, we have

$$\log_2(q_{ij}) = \beta_0 + \beta_1 \times Tsm + \beta_2 \times iPSCTn + \beta_3 \times iPSCTsm \quad (7)$$

Details on how to estimate q_{ij} can be found in Love, Huber, and Anders 2014⁸ and Anders and Huber 2010⁷.

For each gene, our design matrix will have 12 rows representing the 12 samples, and 4 columns, with the first column as the intercept, which represents our baseline cell category Tn, and the next three columns representing the three cell categories Tsm, iPSCTn, and iPSCTsm.

3.2.4 Hypothesis Test for Differential Expression

Using the function `DESeq()`, we perform the Wald test and calculate their corresponding p -values for each gene. P -values are adjusted to control the false discovery rate by the method of Benjamini and Hochberg. A gene is determined to be differentially expressed if the adjusted p -value for the estimated parameter value is smaller than 0.05. We performed the following three hypothesis tests for each gene:

$$\beta_1 = 0$$

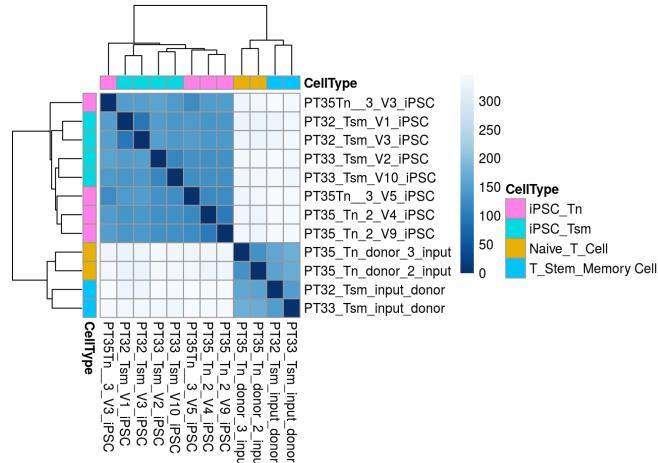
$$\beta_2 = 0$$

$$\beta_1 = \beta_3$$

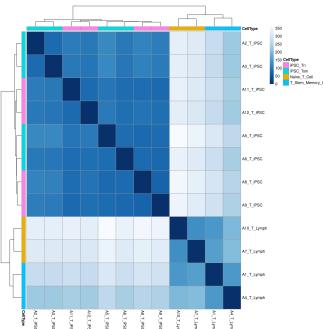
4 Results

4.1 Heat Maps of ATAC- and RNA-seq Samples

Datasets for RNA- and ATAC-seq were first normalized using methods described in Section 3.2.2. Figure 2a and 2b show heatmaps clustering the 12 samples for RNA- and ATAC-seq, respectively.



(a) Heatmap of Unnormalized Counts of RNA-seq across 12 samples



(b) Heatmap of Unnormalized Counts of ATAC-seq across 12 samples

Figure 2. PCA for RNA- and ATAC-seq across 12 samples

In the heatmaps for both ATAC- and RNA-seq, we see clear clustering between the T lymphocyte cells and reprogrammed iPSC cells. Hierarchical clustering demonstrates differentiation between naive and stem memory T cells. Substructure within the iPSC celltype shows a potential iPSC Tn cell in the RNA-seq. ATAC-seq clustering for the iPSC samples show strong clustering between donors.

4.2 PCA of ATAC- and RNA-seq Samples

Next, we ran PCA on both ATAC-seq and RNA-seq datasets, using the normalization methods from DESeq2 and edgeR. PCA results are shown in Figure 3. The first principle component indicates differentiation between T lymphocytes and iPSC cells. The second principle component indicates differentiation between the naive and stem memory subgroups. However, the second principle component confounds differentiation between the iPSC naive and stem memory cells. As we will see in the following sections, lack of differentiation between naive and stem memory iPSC cells affects further analysis of differential expression.

4.3 Comparison between edgeR and DESeq2

As mentioned above, both edgeR and DESeq2 were used for differential expression analysis. In this section we compare the two methods by looking at their estimated log fold changes for both RNA-seq and ATAC-seq.

We see that overall most genes/chromatin open regions selected by the two methods agree well, i.e. estimated log fold changes lie on a diagonal line. While there are some deviance from the diagonal in Tn and Tsm cell comparisons, note that this pair is not really pertinent to our hypotheses and therefore would not pose a huge problem in downstream analysis.

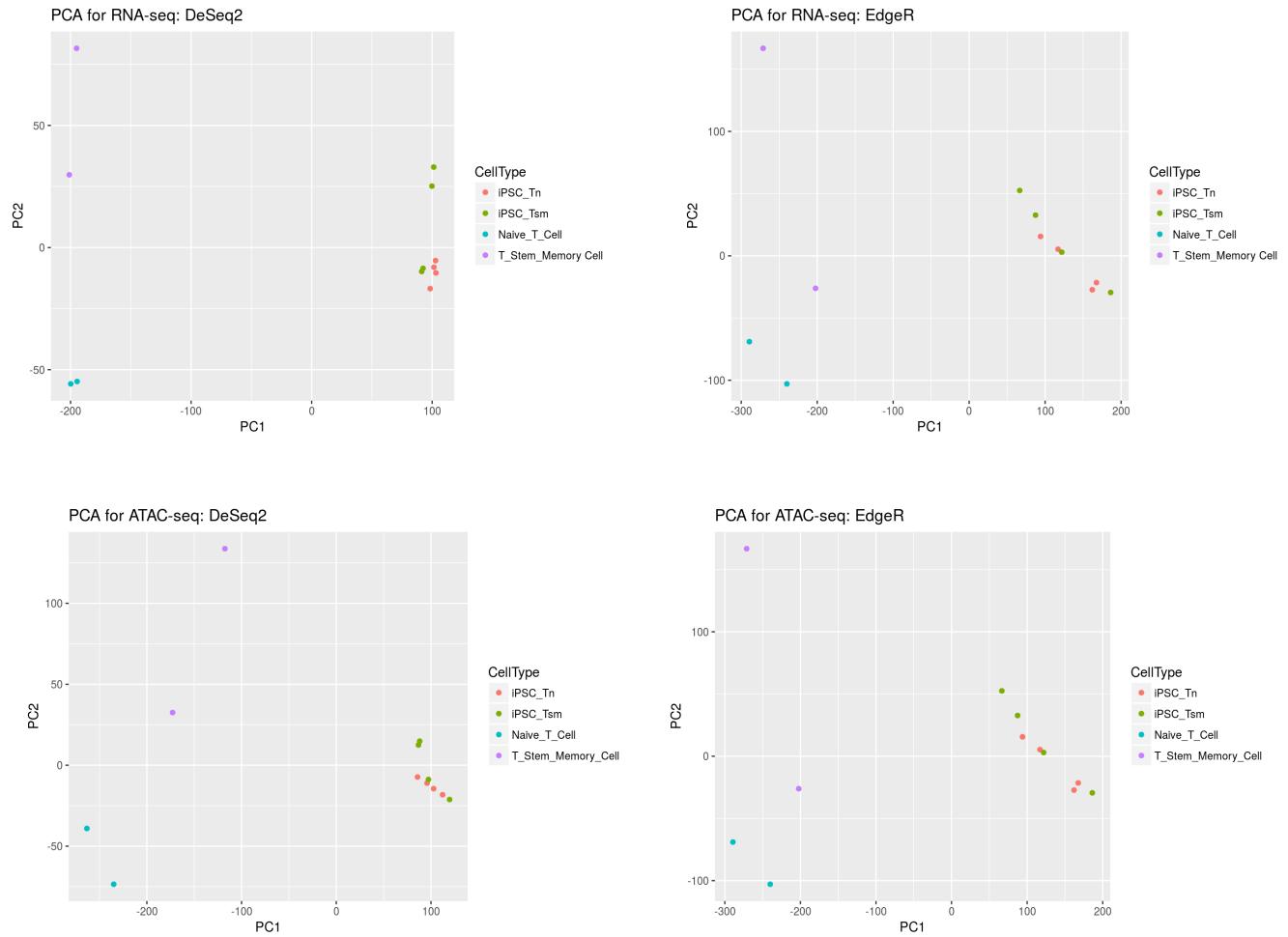


Figure 3. PCA Plots for RNA-seq and ATAC-seq datasets. Normalization techniques from DESeq2 (left) and edgeR (right). Top row shows RNA-seq, bottom row shows ATAC-seq results.

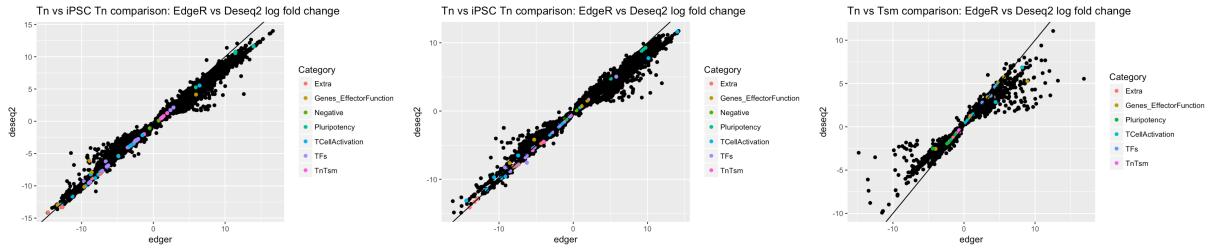


Figure 4. Plot of estimated log fold change of RNA-seq datasets using edgeR and DESeq2. The genes have been annotated with functions.

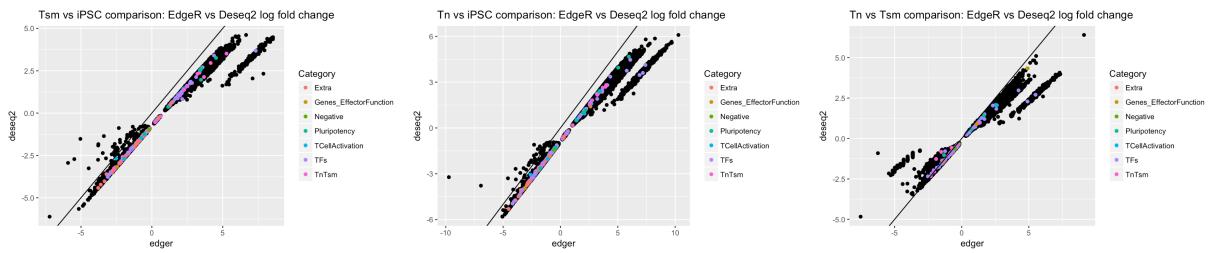


Figure 5. Plot of estimated log fold change of ATAC-seq datasets using edgeR and DESeq2. The genes have been annotated with functions.

4.4 Differential expression analysis on ATAC-seq

In this section we look at volcano plots of ATAC-seq data using both edgeR and DESeq2 to understand the chromatin regions accessibility that have both large fold change as well as high statistical significance.

4.4.1 Tn versus iPSC Tn

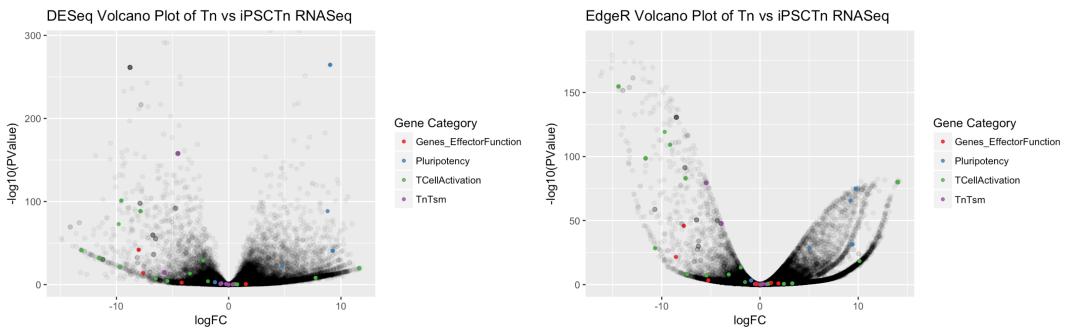


Figure 6. Volcanop plots of ATAC-seq datasets using edgeR and DESeq2 for Tn versus iPSC Tn. The genes have been annotated with functions.

Overall compared to Tn cells, iPSC Tn cells display a higher accessibility for regions associated with pluripotency associated genes and a lowered level of accessibility for Tn/Tsm and T cell activation genes. Biologically our differential analysis validates the cellular heterogeneity of the Tn and iPSC Tn cells.

4.4.2 Tsm versus iPSC Tsm

Overall compared to Tsm cells, iPSC Tsm cells, similar to its naive cell counterpart, have more accessible chromatin regions associated with pluripotency and more restricted regions associated with Tn/Tsm and T cell activation genes. The noticeable difference is in genes effector functions which would be of interest to study in the future.

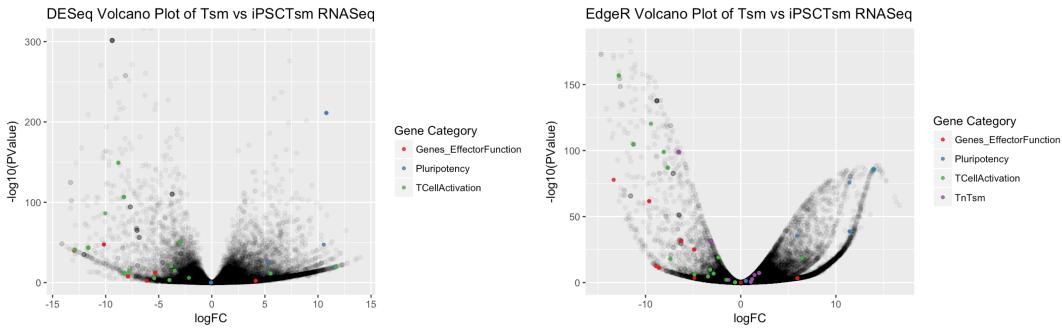


Figure 7. Volcano plots of ATAC-seq datasets using edgeR and DESeq2 for Tsm versus iPSC Tsm. The genes have been annotated with functions.

4.5 Differential expression analysis on RNA-seq

In this section we look at volcano plots of RNA-seq data using both edgeR and DESeq2 to understand the genes that have both large fold change as well as high statistical significance in terms of their mRNA abundance.

4.5.1 Tn versus iPSC Tn

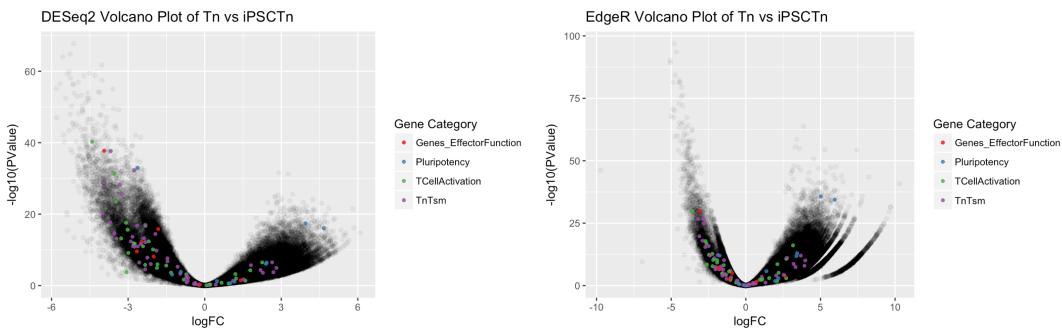


Figure 8. volcano plot of RNA-seq datasets for Tn versus iPSC Tn cells using edgeR and DESeq2. The genes have been annotated with functions.

In this section we look at volcano plots of RNA-seq data using both edgeR and DESeq2 to understand the genes that have both large fold change as well as high statistical significance in terms of their mRNA abundance for Tsm versus iPSC Tsm cells.

4.5.2 Tsm versus iPSC Tsm

In addition to the general trend of Tn versus iPSC Tn cells comparison, Tsm versus iPSC Tsm seems to preserve the Tn/Tsm marker genes better as the expression level of those marker genes distributes more evenly over over-expression and under-expression⁹.

4.6 Integration of ATAC-seq and RNA-seq results

4.6.1 Overview of integration of Seq data

To determine whether cell type specific open chromatin regions from the ATAC-seq analysis correlates with the cell type specific gene expression from the RNA-seq analysis, we integrated our ATAC-seq and RNA-seq data by merging on the gene symbols and results were visualized using Venn Diagram¹⁰. To increase the robustness of our integration results, we also included two different methods (edgeR and DESeq2) in the data integration. Therefore for each comparison of interest, i.e. Tn versus iPSC Tn or Tsm versus iPSC Tsm, we will have four different sets of differentially expressed genes. A final caveat is that the output for ATAC-seq is chromatin binding region instead of the actual mRNA reads, therefore a mapping of the chromatin region back to the gene symbol was performed using annotation data for known genes from UCSC genome browser.

Overall, in Tn versus iPSC Tn cells, across all four different methods, 5195 genes that were expressed at significantly different levels (defined as FDR < 0.05) had at least one associated iPSC Tn cell specific open chromatin region that was not

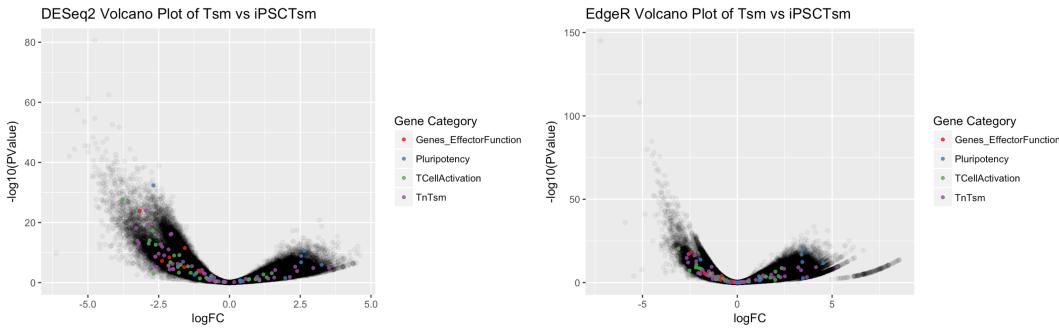


Figure 9. Plot of estimated log fold change of RNA-seq datasets Tsm versus iPSC Tsm cells using edgeR and DESeq2. The genes have been annotated with functions.

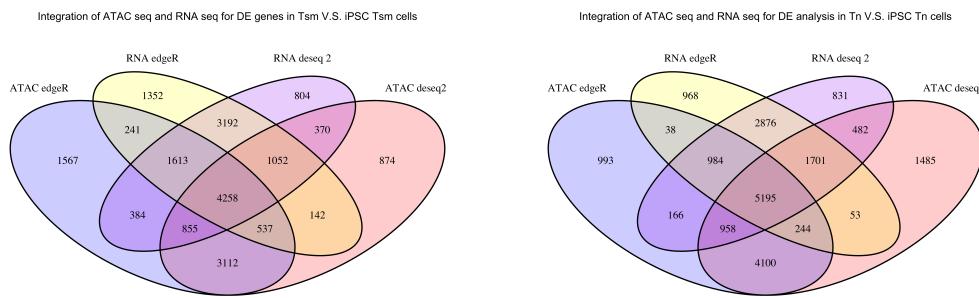


Figure 10. Venn Diagram visualizing integration of ATAC-seq and RNA-seq data on differentially expressed genes using 2 different methods. On the left is for Tsm versus iPSC Tsm cells and the right panel shows integrated data for Tn versus iPSC Tn cells

identified in Tn cells, which accounts for 43% and 39% of all differentially expressed genes depending on whether edgeR or DESeq2 was used for analysis. The corresponding proportion is 40% and 36% for open chromatin regions.

The high proportion of overlapping differentially expressed genes also holds in the Tsm versus iPSC Tsm cells comparison. We see that across all four different methods, 4258 of the genes were both differentially expressed and had at least one associated iPSC Tsm(Tsm) cell specific open chromatin region that was not identified in Tsm(iPSC Tsm) cells. Such genes make up 34% and 33% of the RNA seq and ATAC seq data using DESeq2.

4.6.2 Analysis of Seq data using DESeq2

To further our analysis on the integration, we want to restrict ourselves to RNA-seq versus ATAC-seq using DESeq2 as the main analysis tool. We have produced below the similar Venn diagram for only two of them to render our data more interpretable. The following plot is the integration of ATAC-seq and RNA-seq data on differentially expressed genes using DESeq2 only:

We see that, the proportion of overlap is 63.28% of RNA-seq genes in Tn versus iPSC Tn and 52.16% in that of Tsm versus iPSC Tsm. These results suggest that open chromatin may be a better predictor of gene activation in Tn and iPSC Tn cells than in Tsm and iPSC Tsm cells, perhaps due to inherent differences in gene regulation in these two different cell populations and cellular heterogeneity. When interpreting the results, one should also remember that there is potential loss of information when mapping the region of chromatins to the gene symbols as on average about 80% of the differentially ATAC seq peaks were mapped to genes.

4.6.3 Distribution of adjusted p values

Finally, the previous trend was also captured by the plot of adjusted p values: we see that ATAC-seq data in general exhibits more higher degree of differential expression, i.e. skewed towards the lower end. In addition, it seems that in Tn cell, DESeq2 appears more sensitive to reprogramming differentiation while in Tsm cells edgeR is more sensitive to reprogramming differentiation. It is definitely an interesting direction to follow up.

Overall, we see that the integration of RNA seq and ATAC seq established the high degree of differential expression in both Tn versus iPSC Tn and Tsm versus iPSC Tm cells. This refutes the notion of "preservation of genetic expression" of Tn and

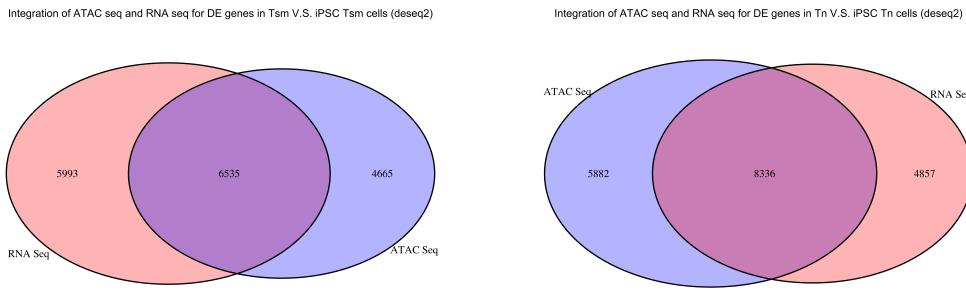


Figure 11. Venn Diagram visualizing integration of ATAC-seq and RNA-seq data on differentially expressed genes. On the left is for Tsm versus iPSC Tsm cells and the right panel shows integrated data for Tn versus iPSC Tn cells

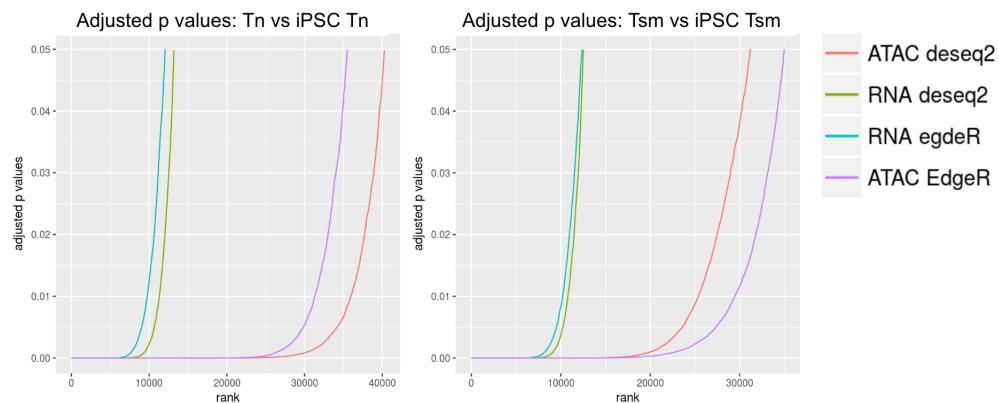


Figure 12. Integration of ATAC-seq and RNA-seq data on differentially expressed genes in Tsm versus iPSC Tsm cells

Tsm cells after reprogramming into their iPSC counterparts. In addition we saw relatively high agreement between edgeR and DESeq2 in terms of differential expression analysis.

5 Discussion

From analysis of reprogramming T cells, it has been observed that much of the epigenetic and transcriptomic landscapes are degraded after reprogramming. Initial PCA results shown in Figure 3 show that although the first principle component differentiates T and iPSC cells for ATAC- and RNA-seq datasets, the second principle component does not differentiate iPSC naive and stem memory cells. Through further analysis of volcano plots after differentiation with edgeR in Figure 9, pluripotent cells are upregulated in the reprogrammed cells for both naive and stem memory, and T cell genes related to T cell activation function is mostly degraded.

5.1 Analysis of Transcription Factors in Reprogramming

We analyzed the change in differentiation for transcription factors Prdm1 and Bcl6. Prdm1 is known to be up-regulated in T cell differentiation states after stem memory, increasing expression until cell death¹¹. Alternatively, Bcl6 is down-regulated in subsequent differentiation stages, decreasing until cell death.

Prdm1 counts for RNA- and ATAC-seq sets across the four cell types is shown in Figure 13. After reprogramming of stem memory cells, expression and accessibility is lost. This aspect may potentially generate more sustainable Tsm cells after re-differentiation.

Bcl6 counts for RNA- and ATAC-seq across cell types is shown in Figure 14. Expression of Bcl6 decreases after Tsm reprogramming, potentially creating instability after re-differentiation. From these results, it is inconclusive of a general effect of reprogramming on transcription factors crucial to Tsm stability.

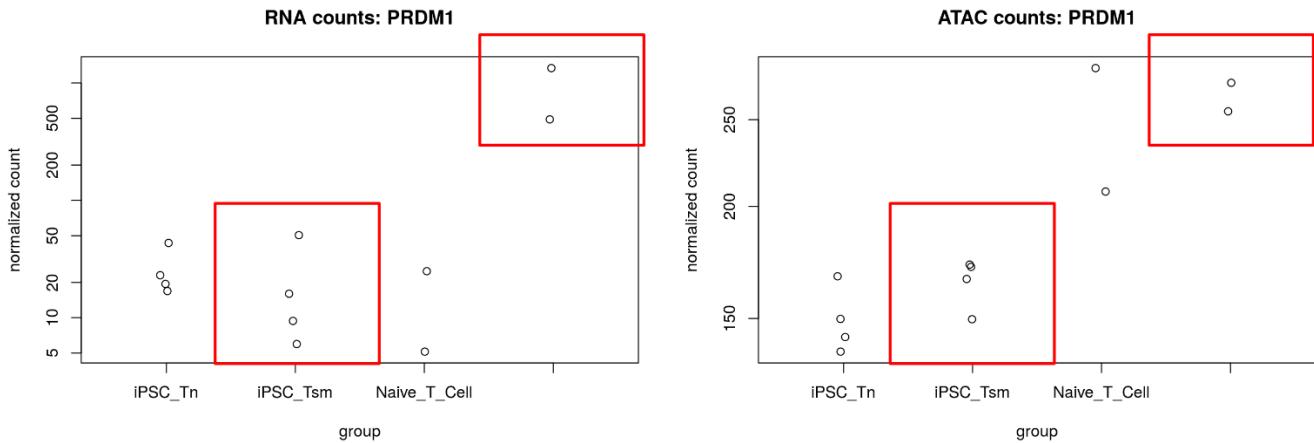


Figure 13. Prdm1 shows loss of expression and accessibility after reprogramming for stem memory cells.

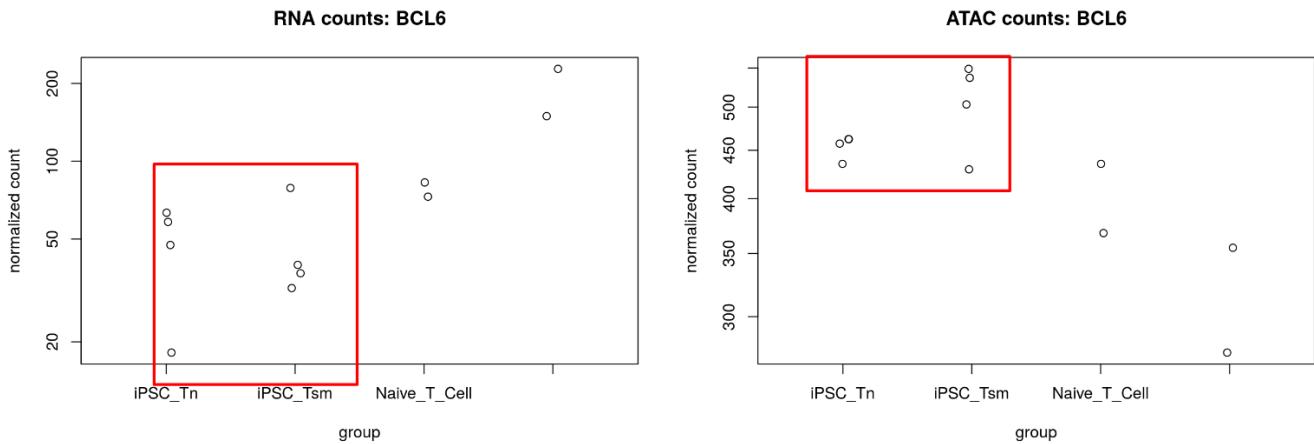


Figure 14. Prdm1 shows loss of expression after reprogramming for both stem memory cells, but increased accessibility exposing the gene.

6 Conclusion and Future Work

In this paper, we investigate differential expression and differential accessibility across T cells reprogrammed to iPSC cells. We find that through reprogramming, a majority of the epigenetic landscape is lost, showing levels of increased pluripotent-specific expression and decreased T cell specific expression in the reprogrammed cells.

6.1 Future Work

For future work, we intend to continue this comprehensive study of differentiation between T lymphocytes and reprogrammed cells. Further analysis includes the following:

- Analysis of ATAC-seq Enhancer Regions** Throughout this paper, we compare ATAC-seq regions that overlap known genes. However, we do not consider analysis of enhancer or promoter regions of the ATAC-seq data in our comparison to RNA-seq differentiation. A more comprehensive analysis of the ATAC-seq dataset would consider a union of both enhancer and overlapping gene regions.
- Analysis of Transcription Factors Driving Reprogramming** Thus far, our methods have relied on Yamanaka factors for reprogramming T lymphocytes. However, other transcription factors driving reprogramming have not been identified. By comparing the union of differentially expressed transcription factors after reprogramming, we can identify factors crucial to differentiation.

However, further analysis is constrained by the available datasets. Next steps for data analysis includes the following:

1. **Collection of iPSC cells reprogrammed from fibroblasts** With the collection of iPSC cells reprogrammed from a cell other than T cells, we can compare the epigenetic landscape of the different reprogrammed pluripotent cells. Through this analysis, we can discover whether maintained expression is specific T cell reprogramming.
2. **Differentiation of iPSC Tsm cells to Tsm** After reprogramming, we would like to investigate re-differentiation of iPSC cells back to Tsm cells. Although much of the epigenetic landscape is lost on reprogramming, assessing differential expression in the differentiated iPSC Tsm and Tsm cells would give us an idea of the accuracy of the end to end reprogramming process.

References

1. Brentjens, R. *et al.* Cd19-targeted t cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci. translational medicine* **5** (2013).
2. Klebanoff, C. *et al.* Central memory self/tumor-reactive cd8+ t cells confer superior antitumor immunity compared with effector memory t cells. In *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, 9571–6 (2005).
3. Wherry, E. T cell exhaustion. *Nat. immunology* **12**, 492–9 (2011).
4. Barrett, D. *et al.* Relation of clinical culture method to t-cell memory status and efficacy in xenograft models of adoptive immunotherapy. *Cytotherapy* **16**, 619–30 (2014).
5. Jensen, M. & Riddell, S. Design and implementation of adoptive therapy with chimeric antigen receptor-modified t cells. *Immunol Rev* **257** (2014).
6. Robinson, M., McCarthy, D. & Smyth, G. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma.* **26** (2010).
7. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* (2010).
8. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol.* **15**, 550 (2014). URL <http://dx.doi.org/10.1186/s13059-014-0550-8>. DOI 10.1186/s13059-014-0550-8.
9. Roberto, A. e. a. Role of naive-derived t memory stem cells in t-cell reconstitution following allogeneic transplantation. *Blood* **125**, 2855–2864 (2015). DOI <https://doi.org/10.1182/blood-2014-11-608406>.
10. Ackermann AM Wang Z, N. A. K. K., Schug J. Integration of atac-seq and rna-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016). DOI 10.1016/j.molmet.2016.01.002.
11. Shin, H. *et al.* A role for the transcriptional repressor blimp-1 in cd8(+) t cell exhaustion during chronic viral infection. *Immun.* **2**, 309–20 (2009).

Acknowledgements

We acknowledge our collaborators Bruce Blazar, Nick Haining, and Dharmeshkumar Patel in data collection and aid in data analysis.