Application Note

strandngs

Streamlining NGS Data Management & Analysis

Integrative RNA and ChIP-Seq analysis of regulatory T-cells in Strand NGS

Aishwarya Narayanan, Veena Hedatale and Rohit Gupta

Strand Life Sciences

www.strand-ngs.com

Analyze | Visualize | Annotate | Discover

strand
New Generation Healthcare

Application Note

Integrative RNA and ChIP-Seq analysis of regulatory T-cells in Strand NGS

Aishwarya Narayanan, Veena Hedatale and Rohit Gupta*

Strand Life Sciences

March 2015

* *Corresponding Author (rohit@strandls.com)*

Strand NGS supports multiple workflows like DNA-Seq, ChIP-Seq, RNA/small RNA Seq and Methyl-Seq along with an ability to do integrative analysis between these different workflows. In this application note, we will look at an integrative study involving RNA-Seq and ChIP-Seq data sets.

1 Introduction

Many next-generation sequencing studies utilise multiple technologies to answer complex biological questions. For example a DNA-Seq study elucidating genes with deleterious mutations may lead to a RNA-Seq study to observe and correlate the expression profiles of those genes. Similarly changes in small RNA expression derived from a small RNA-Seq study may be correlated with a RNA-Seq study to identify genes showing negative correlation with their corresponding regulatory small RNAs. So it becomes very important for a NGS analysis software to be able to handle data from multiple technologies as well as enable the user to perform integrative analysis on multiple data sets that may offer complementary information.

For this case study, we followed a paper titled "Next-generation insights into regulatory T-cells: expression profiling and FoxP3 occupancy in Human" [1] to illustrate how one can do an integrative data analysis in Strand NGS. This paper focuses on two types of immune cells: T regulatory cells and T helper cells. Regulatory T-cells (Treg) play an essential role in the negative regulation of immune answers and the transcription factor (TF) FoxP3 is responsible for the regulation of many genes involved in the Treg gene signature. Therefore, a deleterious event in FoxP3 may lead to severe immune deficiencies in human and mice. This paper analysed the RNA-Seq data of both cell types in resting as well as activated state. Further ChIP-Seq data is also analysed for both cell types in activated state. Refer to Figure 1 for study design.

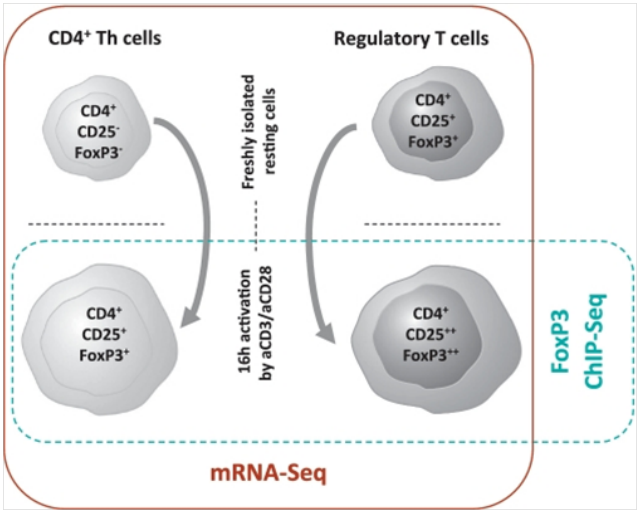

Figure 1: *Activated and Resting states of Treg and Th cells (taken from [1])*

Using RNA-Seq data analysis workflow in Strand NGS, we will first quantify the changes in the transcriptional profile of Treg and Th cell types in resting versus activated state. Further we also observe the differences in gene signatures between the two cell types in both resting and activated states. Using ChIP-Seq workflow in Strand NGS, we will identify the DNA-protein binding sites of FoxP3 in the activated Treg and Th cell types.

Finally, in the integrative part of the analysis, we will correlate the binding sites of FoxP3 as observed by ChIP-Seq to the expression profiles observed via RNA-Seq approach. This analysis will provide insights regarding the impact of FoxP3 binding on regulating the genes in both Treg and Th cell types.

2 Datasets

2.1 RNA-Seq

The data consists of naturally occurring human CD4+CD25+ regulatory T-cells (Treg) and untouched human CD4+ T helper cells (CD4+ Th) along with cells on which polyclonal T-cell activation was done using anti-CD3 (clone OKT3) antibody and anti-CD28 (clone 28.2) antibody. This amounts to four samples of resting and activated primary CD4+ Th and Treg cells. The data is paired end and was generated using the Illumina Genome Analyzer platform and consists of 35bp length reads. Refer to table 1 for more details.

Sample	Cell line	Total reads
SRR192425	T-reg resting	13,655,820
SRR192534	T-reg activated	15,277,248
SRR192532	CD4+ Th resting	8,237,652
SRR192536	CD4 + Th activated	10,335,712

Table 1: *Description of RNA-Seq data sets.*

2.2 ChIP-Seq

Samples corresponding to the ChIP assays of the activated CD4+ Th and T-reg cells were available with two biological replicates (Donors A1/A2 and B1/B2) with each donor in turn having two technical replicates.The data for ChIP-Seq is only from the activated cells i.e. naturally occurring human CD4+CD25+ regulatory T-cells (Treg) and untouched human CD4+ T helper cells (CD4+ Th) on which polyclonal T-cell activation was done for 16h using anti-CD3 (clone OKT3) antibody and anti-CD28 (clone 28.2) antibody. Post genomic DNA purification, FoxP3-bound genomic DNA regions were isolated using a goat polyclonal antibody against FoxP3. The data is single end and was generated using the Illumina Genome Analyzer platform and consists of 35bp length reads. Refer to table 2.

Sample	Donor	Cell line	Total reads
SRR192542	Donor B1	T-reg activated	7,056,313
SRR192543	Donor B1	T-reg activated	7,170,203
SRR192544	Donor B2	T-reg activated	7,507,454
SRR192545	Donor B2	T-reg activated	7,889,442
SRR196006	Donor A1	CD4+ Th activated	6,891,095
SRR192539	Donor A1	CD4+ Th activated	7,267,442
SRR192540	Donor A2	CD4+ Th activated	7,056,313
SRR192541	Donor A2	CD4+ Th activated	7,170,203

Table 2: *Description of ChIP-Seq data sets.*

Because of labelling issues, samples B1 and A2 were discarded. The data can be accessed at http://sra.dnanexus.com/studies/SRP006674/runs.

3 Methods

3.1 RNA-Seq

Alignment was done using our in-house Strand NGS aligner which follows the Burrows-Wheeler Transform (BWT) approach. The data was aligned against both the transcriptome (using Ensembl transcript model) and the genome(hg19). This option allows for reads to map to both the transcriptome and the genome depending on where they align best. The alignment parameters used allowed for 10% of mismatches and 5% of gaps in a read. Reads aligning to multiple places were reported only once and reads aligning to more than 5 places were ignored. Since the base quality dipped towards the 3 end, we also trimmed low quality bases ($\leq$ 10) from that end. To ensure that this trimming did not result in very short reads, we fixed the minimum read length at 25.

Post alignment, the quantification was done based on the methods suggested by Mortazavi at al[7]. As the alignment was done using the Ensembl genes and transcripts model (57751 total genes), the quantification was also done using the same. After quantification, we normalized the data using the RPKM method. Before proceeding to find out the differentially expressed genes, we decided to filter out genes with very few reads ($\geq$ 8 RPKM) in all 4 samples. On these list of genes, fold change (FC) was done across different conditions and the resulting genes (FC$\geq$2.0) were used for downstream pathway analysis. The pathway analysis module computes the $p-value$ using the hypergeometric distribution and pathways satisfying a cutoff of $\leq$ 0.05 were retained. This significance analysis was carried on the curated immune system pathways from Reactome [3].

3.2 ChIP-Seq

The raw reads corresponding to ChIP-seq were aligned in Strand NGS against the human genome (hg19) reference. The raw reads were aligned with minimum of 90% identity; maximum of 5% gaps and 25bp as the minimum aligned read length.

Post alignment, peak were detected with both the MACS[10] and PICS[9] algorithms. Significant peaks were detected with the PICS algorithm in each of the treatment samples without a control, using a window size of 250 bases and a maximum number of 5 binding sites per gene. Peak detection in MACS was carried out in each of the treatment samples without a control, using an average fragment size of 300 bases and a $p-value$ cut off of $1.0E-5$. The resulting peak regions were both annotated with genes present in a $+/-5kb$ window.

After peak calling, motif detection was done on those peak regions overlapping with up and down regulated genes specific to Treg cells and bound by FoxP3. The motif detection was done using the GADEM algorithm[5].

3.3 Correlation between RNA-Seq and ChIP-Seq

Genes resulting from differential expression across conditions in RNA-Seq were compared using the Venn diagram feature in Strand NGS. These differentially expressed genes from RNA-Seq were then compared with the genes resulting from annotating peak calling regions in ChIP-Seq.

Pathway analysis was done on significant genes across both the workflows. The set of pathways and $p - value$ cut-offs used were similar to those in the RNA-Seq study.

4 Results and Discussion

4.1 RNA-Seq

4.1.1 Data QC and Alignment

Before proceeding with the alignment, we looked at some pre-alignment QC metrics to gauge the read quality. In all the samples, most of the reads seem to have an average read quality of around 15 (Refer to figures 2 and 3). This data quality is on the lower side as we typically see Illumina data quality peaking around 30.

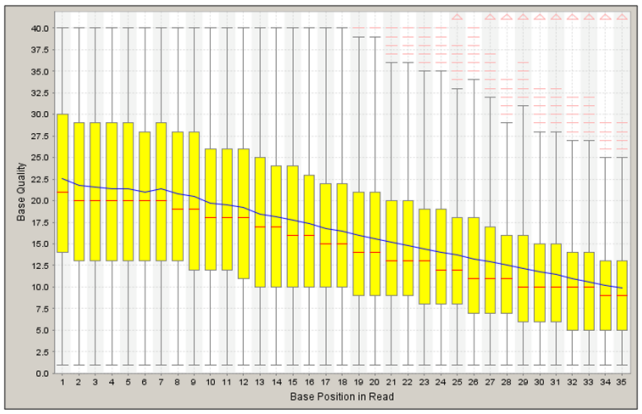

Figure 2: *Base Quality by Position*

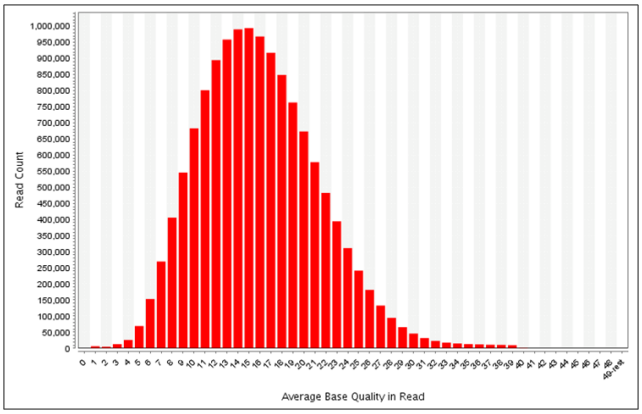

Figure 3: *Read Quality*

After the pre-alignment QC, we decided to proceed with the alignment. A majority of the reads aligned to the transcriptome across all samples (Refer to table3). We can also get further details on the alignment statistics from post alignment QC plots. The alignment score and mapping quality plots shown below give an indication of how many reads aligned with mismatches and mapped to multiple places. Across all samples, a majority of the reads aligned with 0 mismatches (alignment score 100) and mapped to a single location (mapping quality 254). Refer to figures 4 and 5.

Read Information	Rest-Treg	Rest-Th	Act-Treg	Act-Th
Total reads	13,655,820	15,277,248	8,237,652	10,335,712
Aligned reads	11,601,040 (85%)	13,085,460 (85.7%)	6,715,325 (81.5%)	8,423,703 (81.5%)
Aligned to transcriptome	10,106,825	11,414,080	6,042,301	7,344,314
Aligned to genome	1,494,215	1,671,380	673,024	1,079,389

Table 3: *RNA-Seq Read Mapping Statistics*

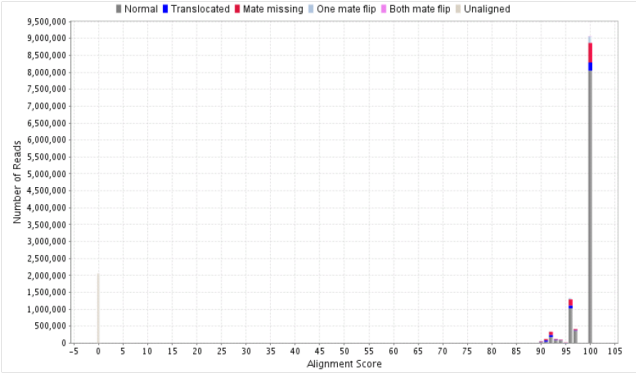

Figure 4: *Alignment Score*

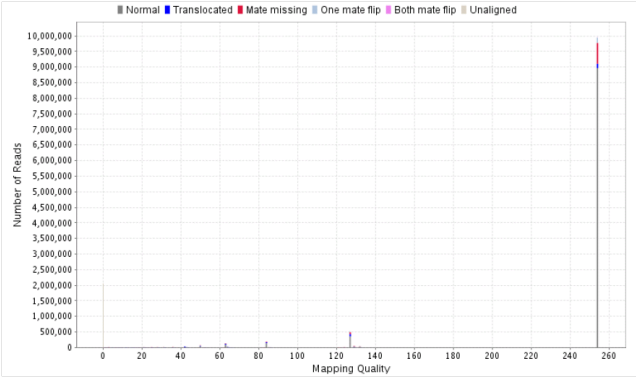

Figure 5: *Mapping Quality*

4.1.2 Data Filtering and Fold Change Analysis

Post quantification and filtering, only 7,439 out of 57,751 genes remained and these were used for subsequent differential expression analysis. In the FC analysis, we found 2,546 genes regulated upon activation in Treg of which 1,274 are upregulated and 1,272 are down regulated. A similar comparison in the Th cells revealed 2,151 genes to be regulated upon activation of which 1,174 were upregulated and 977 genes were down regulated. When we compared these lists (Refer to figure 6), we found 1,399 genes that were commonly regulated upon activation in both the activated cells (for eg. SCD, IL22, FoxP3 or TSPAN2), 1,147 genes unique to Treg (for eg. MS4A3, IL10, LRRC32, CCNL2) and 752 genes unique to Th cells (for eg. FOS, NR4A2, TNFSF13b, DACT1). Refer to figures 7, 8, 9.

The results obtained here were compared against the gene lists provided in the publication for similar conditions.

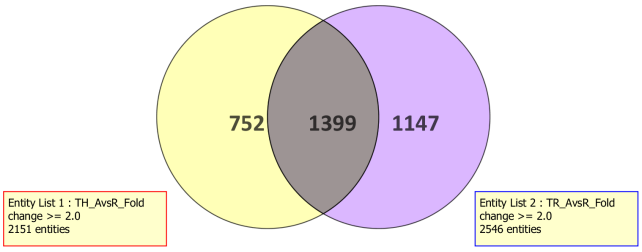

Figure 6: *Common genes between activated Th and Treg cells*

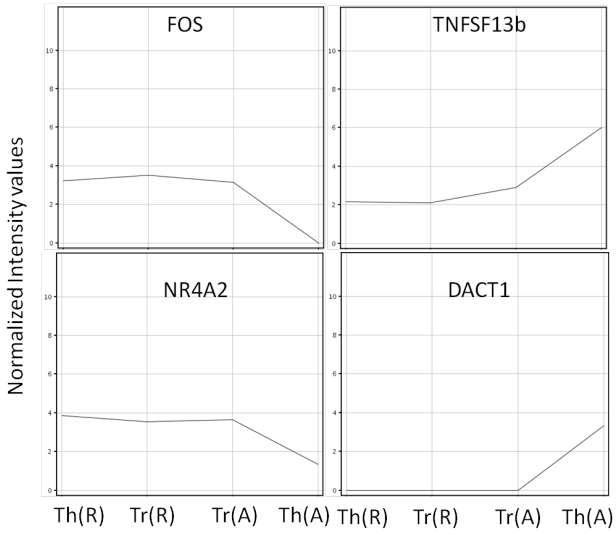

Figure 7: *Genes specifically regulated in Th(Act)*

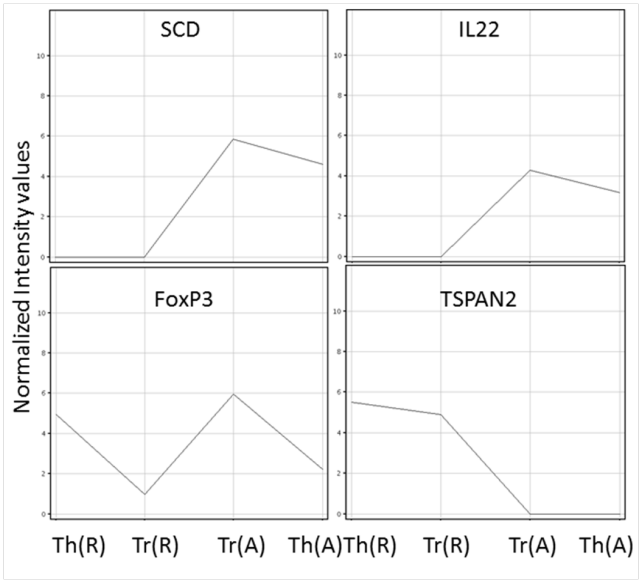

Figure 8: *Commonly regulated in Treg and Th*

Refer to figures 10,11, 12. The differences observed could be attributed to the differences in the methods used. The paper used a FC cut-off of 1.4 along with a SAGEBetaBin significance score of $\leq$ 0.01. We tried applying DESeq to the sample comparisons. But without replicates, the statistical test might not be as powerful and the numbers

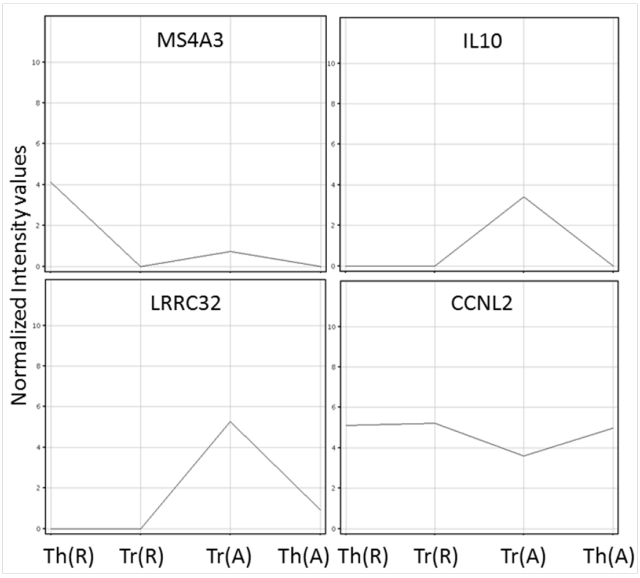

Figure 9: *Genes specifically regulated in Treg(Act)*

reported were on the lower side.

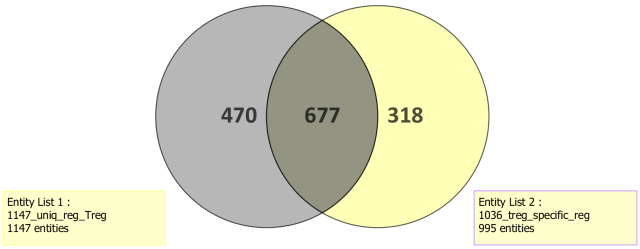

Figure 10: *Comparison of Treg(Act) specific gene signatures*

Figure 11: *Comparison of Th(Act) specific gene signatures*

Figure 12: *Comparison of common gene signatures*

4.1.3 Pathway Analysis

Pathway analysis was done on a set of 1147 genes that were uniquely regulated in Treg, 752 genes that were uniquely regulated in Th and on a commonly regulated gene list of 1399 genes. This yielded 37, 39 and 33 pathways respectively of which 22 are common. These include TCR signalling, TRF mediated TLR3/TLR4 signalling, Toll like receptor cascades pathways. Upon activation, Th cells seem to modulate IPAF and AIM2 inflammasome pathways, IRAK2 and TRAF6 mediated activation of TAK1 complex, the ER phagosome pathway etc. In the case of Treg cells, traffic and processing of endosomal TLR, NOD1/2 signaling pathway, CD28 co-stimulation, complement cascade pathways were some of the ones that came up as enriched.

4.1.4 Fold Change across different cell types

A comparison was also done across the different cell types in the same state. 436 genes were differentially regulated in Treg resting cells compared to Th resting cells; 288 upregulated and 148 down regulated. In the activated state condition, 1597 genes were regulated out of which 483 were upregulated and 614 were down regulated. Out of these 78 were commonly upregulated and 45 genes were commonly down regulated. Refer to figure 13.

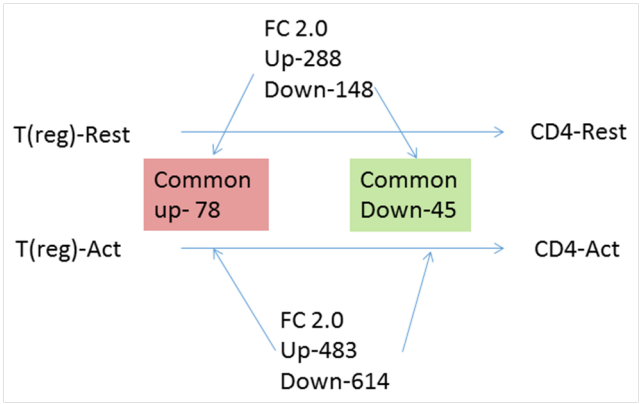

Figure 13: *Comparison across cell types*

4.1.5 Differential Splicing

Along with differential gene expression, the paper also focuses on differential splicing events. One example shown in the paper is the FoxP3 gene. Two different splice variants have been described for FoxP3 in human[8]. One isoform originates from an in-frame exon skipping event of exon 3 (protein coding exon 2) leading to a protein product which lacks 34 amino acids. Neither the forkhead nor the zinc-finger domains are affected by this event. Both variants have also been shown to be functional [8] such that their biological role is still to be elucidated. From the gene view of this event, one can decipher that not only is FoxP3 upregulated in activated Treg compared to other samples but the ENST00000376199 transcript (coloured brown in the read density plot; See figure 14) with the exon skipping event is the transcript that is predominantly expressed in activated Treg.

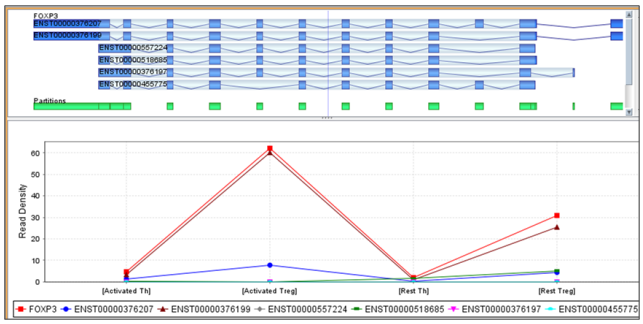

Figure 14: *FoxP3 splicing event in the gene view*

4.2 ChIP-Seq Analysis

4.2.1 Data QC and Alignment

Pre-alignment QC indicates that the majority of the reads have an average base quality of around Q30. The mean base quality by position along the 35 bp read length ranges from Q39 to Q20. Refer to figures15,16.

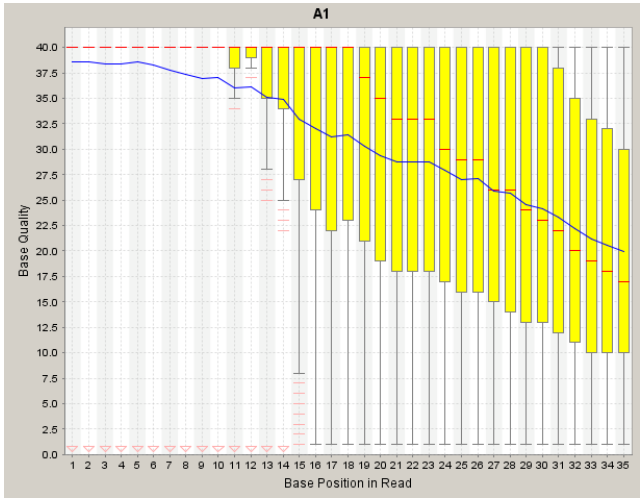

Figure 15: *Base quality by position*

The percentage of aligned reads range from $85-87\%$ for the A1 (CD4+ activated) and B2 (Treg Activated) samples. Refer to table5 for more details. Technical replicates in both the conditions were combined for downstream analysis.

Mapping quality indicates that around 68% reads were uniquely mapped in A1 and around 71% for B2(Mapping quality 254). ENCODE recommends a minimum of 10 million uniquely mapped reads for each biological replicate [4]. The Library Complexity Plot reflects the percentage of uniquely aligned reads in the samples. The library complexity score of 0.73 (range 0 to 1) indicates that the library contains 26.63% duplicate reads (See figure17). When the library has no duplicate reads, the distribution of the library perfectly aligns with the 45 degree line (in blue) and

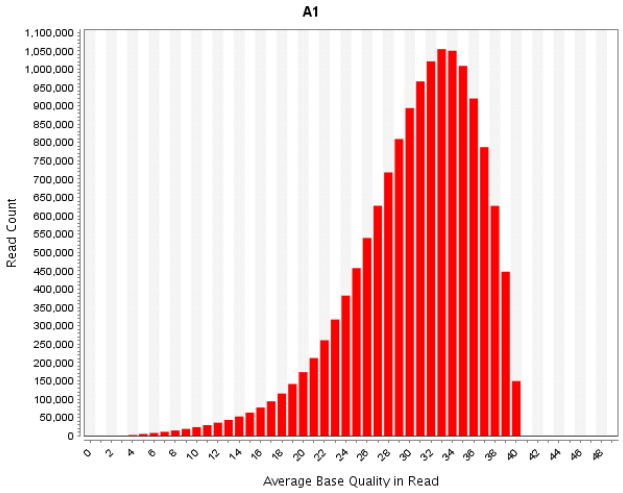

Figure 16: *Read quality*

Sample	Donor	Cell line	Total Reads	Aligned Reads
SRR192544	Donor B2	T-reg activated	7,507,454	6,598,299
SRR192545	Donor B2	T-reg activated	7,889,442	6,873,047
SRR196006	Donor A1	CD4+ Th activated	6,891,095	5,890,341
SRR192539	Donor A1	CD4+ Th activated	7,267,442	6,254,033

Table 4: *ChIP-Seq Read Mapping Statistics*

the library is said to be complex. In addition, the numbers of reads for every region in the positive and negative strand are very similar within a given sample as expected.

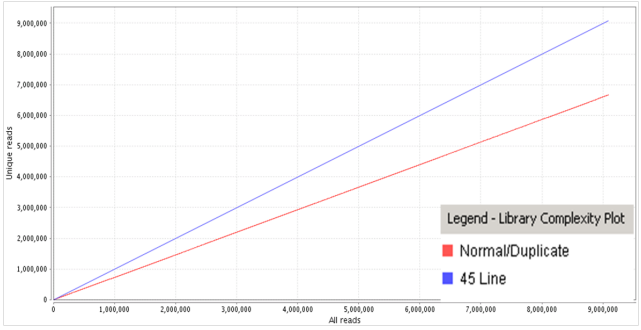

Figure 17: *Library complexity plot to assess duplicates*

4.2.2 Peak Calling

ChIP-seq analysis was performed to identify genomic regions which are bound by FoxP3 in activated CD4+ Th and activated Treg cells. As part of this analysis, peaks were detected with both the MACS[10] and PICS[9] algorithms. The number of peaks detected in the activated Treg sample was 31225 for MACS and 27180 for PICS. These regions corresponded to 13459 genes and 10312 genes ($+/-5kb$ region size). For the Th sample, the results came out to 6083 regions for MACS and 3802 regions for PICS which corresponded to 3296 genes and 1901 genes. (Refer to table5).

Sample	MACS (Regions)	Annotated genes	PICS (Regions)	Annotated genes
Treg	31,225	13,459	27,180	10,312
Th	6,083	3,296	3,802	1,901

Table 5: *Peak Calling Results*

From the above table, we can see that Treg cells show a larger number of peaks compared to the CD4+ Th cells. This is as expected as FoxP3 has a prominent role in Treg cell regulation when compared to Th cells.

4.3 Integrative Analysis

From the 483 genes that came up as upregulated upon activation in Treg when compared to Th(Refer to figure 13), 290 genes were found to be common with the 13459 genes corresponding to the peaks in Treg(MACS). These were the genes that were activated in Treg as well as bounded by FoxP3. A similar comparison of the same 483 geneset with the 3296 genes corresponding to Th peaks(MACS) revealed 74 common genes which were upregulated in activated Treg and also bounded by FoxP3 in Th cells. When we compared these two gene lists, it yielded 71 common genes which were bound by FoxP3 in both cell types and upregulated in Treg. Out of the 219 unique genes bounded by FoxP3 and upregulated in Treg, 29 were found to be also upregulated in the Treg resting state when compared to Th resting state. The 190 genes that were specifically upregulated in Treg upon activation and which were also bound by FoxP3 were used for pathway analysis. 7 pathways came up as significant($p - value \leq 0.05$). These included the downstream signaling events of B cell receptor(BCR), Nucleotide-binding domain, leucine rich repeat containing receptor(NLR) signaling pathways, Innate immune system pathways etc.

An example for a gene specifically upregulated in Treg upon activation and which is also bound by FoxP3 would be the CTLA4 gene which shows a clear up-regulation in activated Treg compared to the other samples in the RNA-Seq study(See figure 18). In the ChIP-Seq study, it shows peaks at two points called out by both PICS and MACS(See figure 19). These were identified as potential FoxP3 binding sites in the paper(See figure 20).

CTLA-4 is a receptor on T cells that plays a critical role in the initial activation and subsequent control of cellular immunity. CTLA-4 is transiently expressed following T cell activation and is a known T-cell marker. The binding of FoxP3 to CTLA4 gene results in increased histone acetylation ([2]). This chromatin modification facilitates gene transcription and serves as a direct activator of CTLA4 gene expression. The signal delivered via CTLA-4 down-regulates T cell function and inhibits excessive expansion of activated T cells.

A similar analysis was done on the down regulated gene list of 614 genes in activated Treg compared to

Figure 18: *RNA-Seq Gene View for CTLA4.*

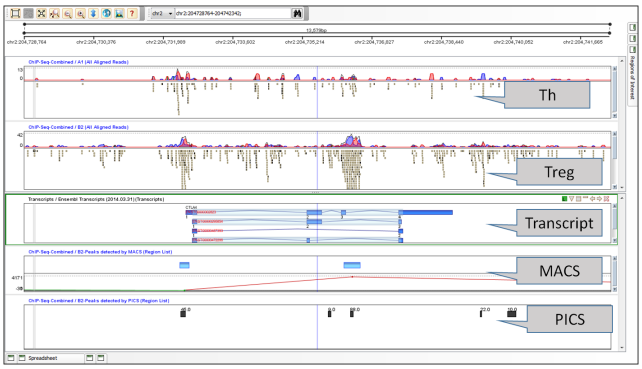

Figure 19: *ChIP-Seq peaks for CTLA4.*

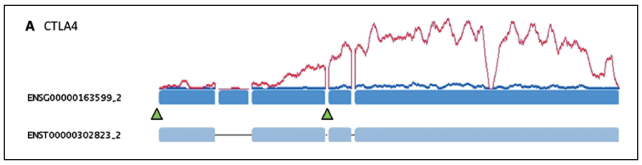

Figure 20: *Binding sites for CTLA4 (taken from [1]).*

activated Th cells(Refer to figure 13). 355 genes were downregulated and bound by FoxP3 in Treg and 99 genes were downregulated in Treg and bound by FoxP3 in Th. These 99 genes were a subset of the 355 genes. 256 unique genes remained that were specifically bounded by FoxP3 in Treg and also downregulated in Treg. 5 of these genes were also downregulated between the resting states of Treg and Th. The remaining 251 genes that were bound by FoxP3 and down regulated in activated Treg cells were used for pathway analysis which resulted in 31 pathways ($p - value \leq 0.05$). These included the TCR signalling pathways, co-stimulation by the CD28 family pathway, TRIF mediated programmed cell death as well as the toll like receptor cascade pathways. These 31 pathways showed a very good concordance with those found in the RNA-Seq analysis.

The peak regions detected by MACS in the activated Treg cells and CD4+ Th cells when compared with promoter regions($+/-5kb$ from gene start site) showed that 3630 regions are in common while 10748 regions are bound by FoxP3 only in Treg cells contributing to a Treg specific regulatory network of FoxP3. The genes corresponding to these regions were later compared with the 190 and 251 up and down regulated genes in activated Treg and bound by FoxP3. (See Figure 21).

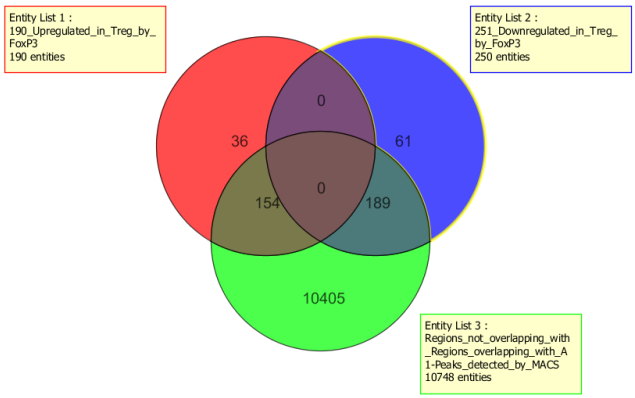

Figure 21: *Comparison of genes bound to promoter regions and up/down regulated genes*

Though a large number of genes seem to be regulated by FoxP3 by binding to promoter regions, a few genes also showed sites after the 5kb boundary from the gene start site. As shown in the case of CTLA4 gene, FoxP3 regulates genes by binding to alternate sites inside the gene. (See figure20). One example gene which was downregulated in activated Treg and bound by FoxP3 but not in the promoter region was CLSTN1. (See figure22).

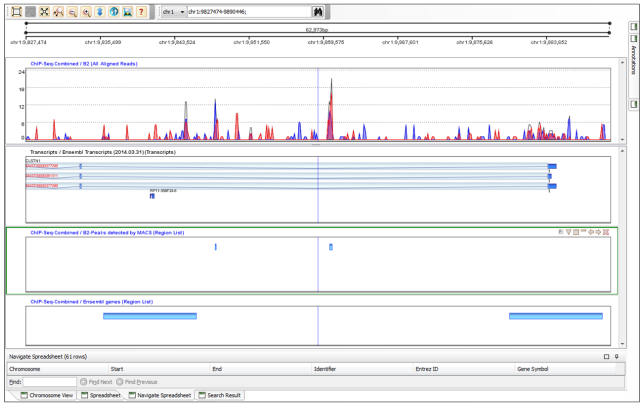

Figure 22: *CLSTN1 binding sites.*

Motifs were detected using GADEM algorithm within MACS peaks that overlapped with the genes upregulated and down-regulated by FoxP3 in activated Treg cells. These motifs were searched in the JASPAR database to find transcription factors that share this motif signature (See figure 23). Regions bound by FoxP3 also show enrichment for other transcription factors such as ETS-1, ZNF263 and IRF1. Although FoxP3 is a key regulator in Treg development, it is becoming increasingly evident that FoxP3 alone only accounts for part of the Treg signature and, for example, the suppression of IL2 and activation of IL2RA in T-cells are also found in FoxP3-deficient Scurfy mice ([6]).

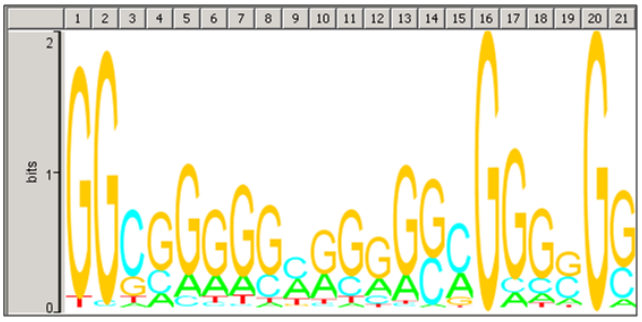

Figure 23: *Motif signature from GADEM*

5 Conclusion

This article demonstrates the applicability of Strand NGS in analysing data in an integrated manner from two next generation sequencing approaches namely RNA-Seq and ChIP-seq. Using RNA-Seq data, we identified gene signatures that are either specific to resting and activated states of Treg and Th cell types or are regulated in both Treg and Th cell types upon activation (Activation signature). Although in some of the previous studies, FoxP3 is shown to have a more suppressive function in organisms other than human, similar to [1], results obtained in this case study indicate that activation and repression of genes by FoxP3 is almost equal in human. Further by using ChIP-seq data, we identified genes that are bounded by FoxP3 in activated Treg and CD4+ Th cells. When correlating the promoter regions with significant FoxP3 peaks, we obtained a set of more than 10,000 genes which may be subjected to a FoxP3 mediated regulation.

Finally we integrated the inferences drawn from individual RNA-Seq and ChIP-Seq data sets to understand the regulatory effects of FoxP3 on the specific gene expression signature of Treg cells. A large number of genes regulated in activated Treg cells compared to activated CD4+ Th cells are bounded by FoxP3 in their promoter regions showing the importance of this transcription factor in the Treg signature. In addition, promoter regions bounded by FoxP3 also showed a enrichment for other transcription factors hinting at a possibility of one or several co-regulatory partners of FoxP3 in regulating Treg genes.

Overall, using the data published in [1], we showed that Strand NGS is a useful tool to not just analyse individual data sets but also very effective in seamlessly integrating the results to understand the correlation across two data sets.

References

[1] Fabian Birzele, Tanja Fauti, Heiko Stahl, Martin C Lenter, Eric Simon, Dagmar Knebel, Andreas Weith, Tobias Hildebrandt, and Detlev Mennerich. Next-generation insights into regulatory t cells: expression profiling and foxp3 occupancy in human. *Nucleic acids research*, page gkr444, 2011.

[2] Chunxia Chen, Emily A Rowell, Rajan M Thomas, Wayne W Hancock, and Andrew D Wells. Transcriptional regulation by foxp3 is associated with direct promoter occupancy and modulation of histone acetylation. *Journal of Biological Chemistry*, 281(48):36828–36834, 2006.

[3] David Croft, Antonio Fabregat Mundo, Robin Haw, Marija Milacic, Joel Weiser, Guanming Wu, Michael Caudy, Phani Garapati, Marc Gillespie, Maulik R Kamdar, et al. The reactome pathway knowledgebase. *Nucleic acids research*, 42(D1):D472–D477, 2014.

[4] Stephen G Landt, Georgi K Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E Bernstein, Peter Bickel, James B Brown, Philip Cayting, et al. Chip-seq guidelines and practices of the encode and modencode consortia. *Genome research*, 22(9):1813–1831, 2012.

[5] Leping Li. Gadem: a genetic algorithm guided formation of spaced dyads coupled with an em algorithm for motif discovery. *Journal of Computational Biology*, 16(2):317–329, 2009.

[6] Wen Lin, Dipica Haribhai, Lance M Relland, Nga Truong, Marc R Carlson, Calvin B Williams, and Talal A Chatila. Regulatory t cell development in the absence of functional foxp3. *Nature immunology*, 8(4):359–368, 2007.

[7] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer, and Barbara Wold. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, 5(7):621–628, 2008.

[8] Emma L Smith, Helene M Finney, Andrew M Nesbitt, Fred Ramsdell, and Martyn K Robinson. Splice variants of human foxp3 are functional inhibitors of human cd4+ t-cell activation. *Immunology*, 119(2):203–211, 2006.

[9] Xuekui Zhang, Gordon Robertson, Martin Krzywinski, Kaida Ning, Arnaud Droit, Steven Jones, and Raphael Gottardo. Pics: Probabilistic inference for chip-seq. *Biometrics*, 67(1):151–163, 2011.

[10] Yong Zhang, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley E Bernstein, Chad Nusbaum, Richard M Myers, Myles Brown, Wei Li, et al. Model-based analysis of chip-seq (macs). *Genome Biol*, 9(9):R137, 2008.

strand

New Generation Healthcare

Strand was founded in 2000 by computer science and mathematics professors from India's prestigious Indian Institute of Science who recognized the need to automate and integrate life science data analysis through an algorithmic and computational approach. Strand's segue into the life sciences was through informatics products and services for research biologists, chemists, and toxicologists that combine advanced visualization, predictive systems modeling, data integration and scientific content management - over 2000 research laboratories worldwide (about 30% of global market share) are licensees of Strand's technology products, including leading pharmaceutical and biotechnology companies, research hospitals and academic institutions. With a recent investment by Biomark Capital, Strand has grown its established team to over 200 employees, many with multidisciplinary backgrounds that transcend computation and biology.

Since 2012, Strand has been expanding its focus to include clinical genomics, spanning sequencing, data interpretation, reporting and counseling. Strand operates a 10,000 square foot laboratory space with state-of-the-art clinical genomics capabilities and is also establishing Strand Centers for Genomics and Personalized Medicine in several hospitals around the world to serve as outreach points for genomic counseling. Based on the experience gained from sequencing, analyzing, interpreting and reporting on clinical samples over a wide variety of clinical indications, Strand has developed an end-to-end solution for clinical labs that handles all stagesfrom analysis to reporting. The interpretation and reporting software platform has been designed and developed specifically for the medical professional, ranging from the molecular pathologist to the physician. By enhancing sequence-based diagnostics and clinical genomic data interpretation using a strong foundation of computational, scientific, and medical expertise, Strand is bringing individualized medicine to the world.

For more information about Strand, please visit **www.strandls.com**,
or follow us on twitter @StrandLife.

INDIA
5th Floor, Kirloskar Business Park, Bellary Road, Hebbal, Bangalore 560024

USA
548 Market Street, Suite 82804, San Francisco, CA 94104