Immunological Genome Project
ImmGen.org

# Ultra-low-input RNA-seq (ULI RNA-seq)

After the final sort of 1,000 cells directly into 5µl lysis buffer (TCL Buffer (Qiagen) with 1% 2-Mercaptoethanol), Smart-seq2 libraries were prepared as previously described (Picelli et al., 2013; Picelli et al., 2014) with slight modifications. Briefly, total RNA was captured and purified on RNAClean XP beads (Beckman Coulter). Polyadenylated mRNA was then selected using an anchored oligo(dT) primer (5′–AAGCAGTGGTATCAACGCAGAGTACT30VN-3′) and converted to cDNA via reverse transcription. First strand cDNA was subjected to limited PCR amplification followed by Tn5 transposon based fragmentation using the Nextera XT DNA Library Preparation Kit (Illumina). Samples were then PCR amplified for 12 cycles using barcoded primers such that each sample carries a specific combination of eight base Illumina P5 and P7 barcodes and pooled together prior to sequencing. Smart-seq paired-end sequencing was performed on an Illumina NextSeq500 using 2 x 38bp reads with no further trimming.

**Primary QC on reads.**

Reads were aligned to the mouse genome (GENCODE GRCm38/mm10 primary assembly and gene annotations vM25; https://www.gencodegenes.org/mouse/release_M25.html) with STAR 2.7.3a (https://github.com/alexdobin/STAR/releases). The ribosomal RNA gene annotations were removed from GTF (General Transfer Format) file. The gene-level quantification was calculated by featureCounts (http://subread.sourceforge.net/; version 2.0.0). Raw reads counts tables were normalized by median of ratios method with DESeq2 package from Bioconductor (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) and then converted to GCT and CLS format. Samples with less than 1 million uniquely mapped reads were automatically excluded from normalization to mitigate the effect of poor quality samples on normalized counts.

**ImmGen QC Process**

Samples having fewer than 8,000 genes with over ten reads were removed from the data set to eliminate the effect of samples with potential PCR amplification errors. All samples were also screened for contamination by using known cell type specific transcripts (per ImmGen ULI RNAseq and microarray data). In practice, the acceptable threshold was set at 1 of typical gene expression of contaminant cell types. Retaining such samples can create structure in the data, and/or generate false distances between samples. In addition, biological replicates were analyzed for Pearson correlation to identify poor quality samples and remove them from the data set. Pearson correlation was calculated on transcripts with an average of greater than five reads or below the 99th percentile for number of reads in the dataset to avoid outlier effects. Any replicates that did not exhibit correlation of 0.9 or greater were removed from the data set prior to downstream analysis. Finally, the RNA integrity for all samples were measured by median TIN across mouse housekeeping genes (https://sourceforge.net/projects/rseqc/files/BED/Mouse_Mus_musculus/mm10.HouseKeepingGenes.bed.gz) with RSeQC software (http://rseqc.sourceforge.net/#tin-py; version 2.6.4). Samples with TIN < 45 were removed from the data set prior to downstream analysis.

**REFERENCES**


Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smartseq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 10, 1096-1098.

Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Fulllength RNA-seq from single cells using Smart-seq2. Nat Protoc. 9, 171-181.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7, 562-578.

Trombetta, J.J., Gennert, D., Lu, D., Satija, R., Shalek, A.K., and Regev, A. (2014). Preparation of single-cell RNA-seq libraries for next generation sequencing. Curr. Protoc. Mol Biol.107, 4-17.