JUMPCODE

Improved Single Cell Transcriptional Profiling with CRISPRclean™

Azeem Siddique,[1,5] Gaia Suckow,[1,5] Jay Carey,[3] Phillip Ordoukhanian,[1,5] Keith Brown[5] (keith@jumpcodegenomics.com), Steve Head,[1,5] Stan Nelson,[4,5] Doug Evans,[2] Ali Torkamani,[2] and Tom Alsaigh[2] | 1. The Scripps Research Institute, La Jolla, CA; 2. Scripps Research Translational Institute, Scripps Research Institute, La Jolla, CA; La Jolla, CA; 3. Fulcrum Genomics, Somerville, MA; 4. David Geffen School of Medicine, UCLA, Los Angeles, CA; 5. Jumpcode Genomics, San Diego, CA

Introduction

Transcriptional profiling has been revolutionized with high throughput, massively parallel, single cell RNA sequencing. Microfluidic/microdroplet technologies recently introduced to the market produce complex data sets that allow investigators to deconvolute complex cell mixtures, identify cell types present in healthy and diseased tissues, and create cell type-specific transcriptional signatures. These technologies are dramatically enhancing our ability to identify transcriptional and cellular perturbations driving disease at the individual cell level. However, the techniques are expensive and enable only sparse sampling of RNAs from each cell, with many genes represented by only 1 or 2 sequencing reads. This limit is due partially to the fact that libraries are dominated by an abundance of housekeeping RNAs, which dominate sequencing reads and limit detection of even moderately expressed transcripts that often drive biological differences between cell types. To improve performance of single cell RNA sequencing (scRNA) sequencing, we have developed a CRISPR-based depletion system called CRISPRclean. CRISPRclean is a post-library depletion method that utilizes CRISPR-Cas technology to deplete uninformative molecules from a sequencing library, allowing sequencer capacity to be efficiently used for greater sensitivity and/or reduced cost. Here we apply CRISPRclean to remove 100 overexpressed, housekeeping transcripts that are abundantly expressed in all cell types. The genes consist of ribosomal and mitochondrial protein coding transcripts (mRNA with polyA tails). We treat fully prepared 10x Genomics single cell libraries with CRISPRclean and demonstrate depletion of sequencing reads associated with the target genes and enhanced sensitivity for other genes.

Methods:

We analyzed multiple, published single-cell studies that employed cells from a variety of tissues and different disease states (including neurological disorders, cardiovascular disease, and soft tissue cancer) to determine a candidate gene list for depletion. Our goal was to target genes of little biological interest that are consistently at high abundance in single cell sequence data. Transcript abundance was ranked first by read count and then filtered for a low coefficient of variation (between disease and normal states) of < 1.1. A set of 100 genes were identified as overabundant and invariant across all sample sets. All of the genes targeted for depletion are mitochondrial and ribosomal protein-coding genes.

CRISPR guide RNAs were designed to the 100 target genes using the Jumpcode Genomics guide design pipeline. The guide RNAs were filtered for optimal sequence, secondary structure, *in vitro* cleavage efficiency and for minimal off-target effects. DNA templates containing a phage T7 RNA polymerase promoter upstream of the full length CRISPR single guide RNA sequence were synthesized on a DNA oligo array (Agilent, Santa Clara, CA), PCR amplified and transcribed *in vitro* to produce a pool of guide RNAs for use in the CRISPRclean protocol. Depletion was tested on single-cell RNA-Seq libraries generated via the 10x Genomics Chromium system at the Scripps Research Institute (La Jolla, CA). The samples, for which data is presented here, comprise of matched diseased and normal atherosclerotic carotid artery tissue collected during carotid endarterectomy from three donors. The normal tissue samples contain approximately 3,500 cells and the diseased tissue samples have approximately 11,000 cells. Each 10x Genomics library was sequenced on an Illumina NextSeq instrument with approximately 250 million 2 x 150 PE reads corresponding to ~60K reads per cell for the normal and ~24K reads per cell for the diseased tissue samples (sequencing saturation of ~85% and ~65% for normal and diseased sample types, respectively). The prepared libraries were subsequently depleted with the CRISPRclean scRNA depletion product. The samples were sequenced on the same NextSeq instrument again, for a similar read length and depth as the original (non-depleted) samples.

All sequence data was processed through Cell Ranger, subject to standard mitochondrial and doublet detection algorithms, and partitioned into cell-types using Monocle3. Differential expression was performed using a generalized linear model to control for patient-specific transcriptomic effects.

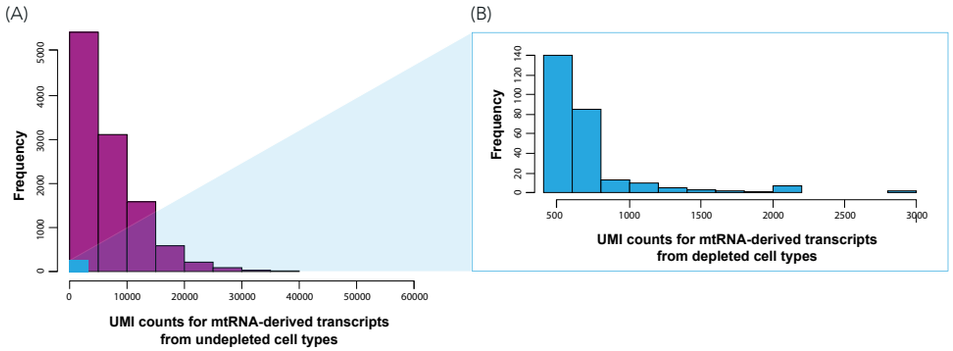

Figure 1: Results plot the number of unique UMIs on the X-axis and the frequency of the individual UMIs from mitochondrial-derived transcripts on the Y-axis. Results show > 95% reduction in total and unique UMI frequency of the depleted library targets.

Genes with altered UMI abundance			
	Increased	Decreased	Unaffected
Healthy tissue	2747	142	18933
Diseased tissue	4260	228	18719

Table 1: The number of genes (relative to total genes for which transcripts are observed in the sequencing data) for which UMI counts are increased, reduced or unaffected as a result of CRISPRclean-mediated depletion.

Results:

Effect of abundant gene depletion on single cell gene expression

Differential gene expression results demonstrate that the vast majority of genes targeted for depletion display the greatest reduction in expression across all genes analyzed in both healthy and diseased samples. Figure 1 shows that mitochondrial gene UMI counts are dramatically lower after depletion. As a result of mitochondrial and ribosomal gene depletion, total UMI counts per cell are lower as well. Importantly, depletion results in an increase of 2,747 UMIs (12.6% of all genes in the library) derived from healthy tissue and 4,260 (18.4% of all genes in the library) derived from diseased tissue (Table 1) indicating that depletion has improved the ability to detect certain genes.

Since depleted libraries are derived from undepleted ones, 10x Genomics cell barcodes are identical between depleted and undepleted libraries of the same type. Consequently, cells that have the same barcodes in the depleted and undepleted libraries can be compared for alterations in UMI numbers per cell. Linear regression plots for example genes like RPS12 and CCL4 are shown in the full app note.

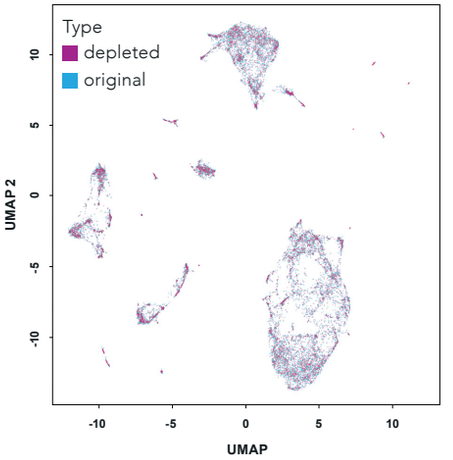

Figure 2: UMAP plot comparing cell type resolution in the original samples (blue) with that of the depleted sample (purple) adjusted to account for the depletion effect on the 100 targeted genes.

Conclusion:

CRISPRclean depletion of abundant and uninformative transcripts increases the ability to detect rare transcripts and adds confidence to differential gene expression data when applied to 10x Genomics single cell RNA sequencing libraries. It may allow the user to glean new biologically relevant information from scRNA-Seq data and, thus, gain novel insight into the genetic and cellular basis for disease.

To view the full application note, visit
jumpcodegenomics.com/CRISPRcleanscRNA