Using CRISPR to Improve Single Cell Transcriptomic Sensitivity

Jon Bezney[1], Dante DeAscanis[1], Sonal Choudhary[1], Jon Armstrong[1], Azeem Siddique[1], Amitabh Pandey[2]

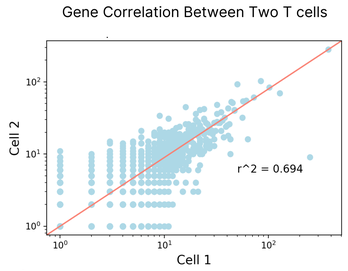
1 - Jumpcode Genomics, 2 - Scripps Research

Design

CRISPR removes wasted reads within scRNA-seq and redistributes them to lowly expressed genes and unique molecules

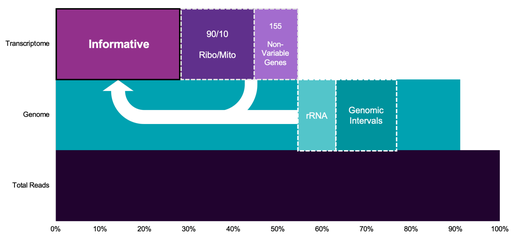
Single cell data is noisy.

Two cells from the same cell type have poor correlation in gene expression. Gene dropout is a pervasive issue that confounds secondary analysis.



A large percent of reads are wasted space.

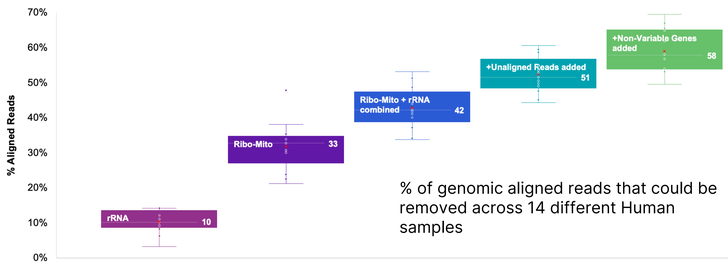
Roughly ~50% of reads that are uninformative for secondary analysis can be removed. The single cell sgRNA target:



- Genomic intervals (non-transcriptome)
- Non polyadenylated RNA
- 90 polyadenylated ribosomal genes
- 10 polyadenylated mitochondrial genes
- 155 non-variable genes

In-silico validation of broad utility.

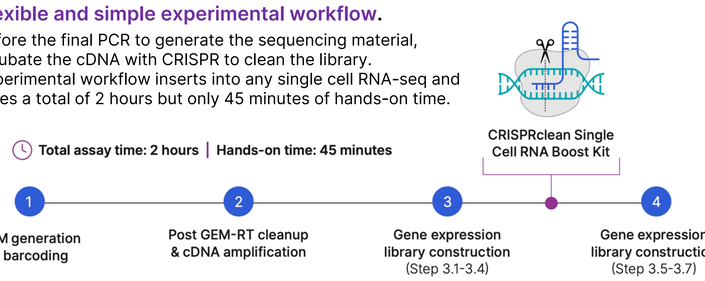
14 different publicly available single cell datasets from various Human biological tissues confirm that 50-60% of reads are un-informative.



% of genomic aligned reads that could be removed across 14 different Human samples

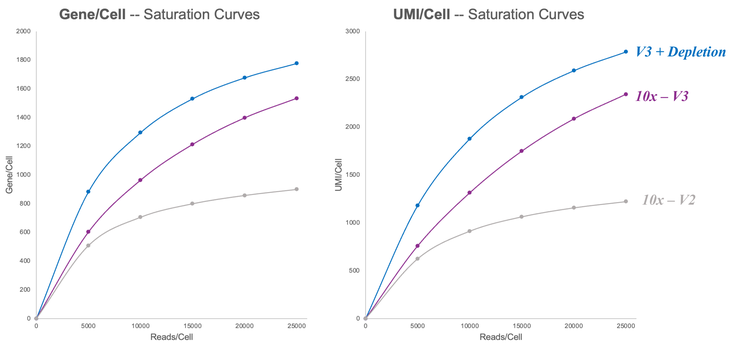
Flexible and simple experimental workflow.

Before the final PCR to generate the sequencing material, incubate the cDNA with CRISPR to clean the library. Experimental workflow inserts into any single cell RNA-seq and takes a total of 2 hours but only 45 minutes of hands-on time.



Random re-distribution improves cell resolution.

Removing highly abundant molecules before PCR enables the capture and identification of lowly expressed genes and unique molecules that otherwise could not have been found. In-silico analysis shows the benefit of performing depletion and comparing the cell sensitivity between 10x Version 2, 10x Version 3, and 10x Version 3 with Jumpcode depletion.

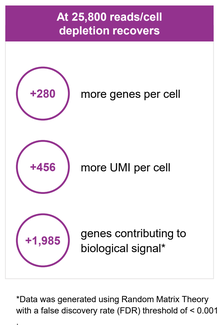


Validation

CRISPRclean was validated on PBMC single cell RNAseq in triplicate (generated using 10x V3 3' gene expression)

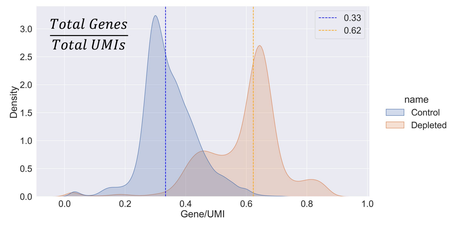
Depletion boosts per cell sensitivity.

Depletion significantly increases the number of genes and UMIs recovered per cell. The control and depleted samples are equivalent splits from the same WTA (cDNA), which means the same exact cells are found in both conditions. For every 1,000 unique molecules found per cell in the control, ~330 genes are identified. For every 1,000 unique molecules found per cell in the depleted, ~620 genes are identified.



*Data was generated using Random Matrix Theory with a false discovery rate (FDR) threshold of < 0.001

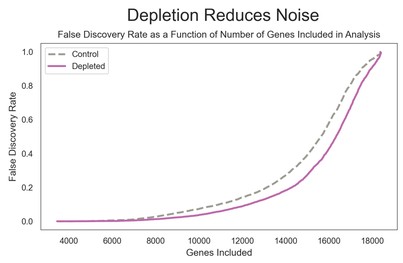
Learning Rate: Total number of genes per cell / total number of UMIs per cell

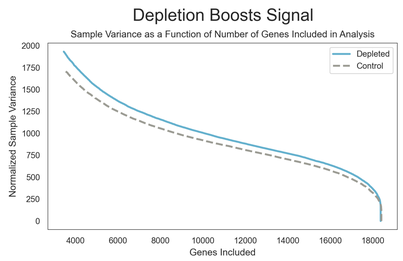


Depletion increases signal to noise.

By reducing the number of drop-outs and redistributing reads to lowly expressed genes and molecules, depletion increases the ability to distinguish between true biological signal and background noise.

Using Random Matrix Theory[1] to distinguish between signal and noise: as the number of genes is gradually increased and included in downstream analysis it becomes apparent that the noise is reduced (false discovery rate is always lower in the depleted condition) while the biological signal is enhanced (sample variance is always higher in the depleted condition).

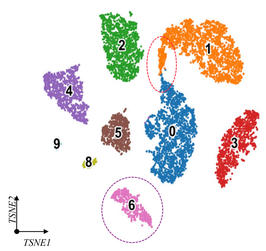




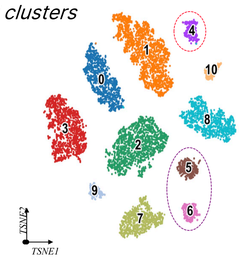
Depletion increases total cell cluster identification.

Using a deep learning unsupervised clustering algorithm (DESC)[2], the depleted condition identifies additional clusters. Despite being an unsupervised method, it is still required by the user to input the clustering resolution (Louvain/Leiden resolution). Various resolutions were chosen from the lower limit to the upper limit and the depleted condition always contained more clusters. The graphs below represent a resolution that produced a total number of clusters typical for PBMC samples.

PBMC Control: *9 clusters*



PBMC Depleted: *11 clusters*



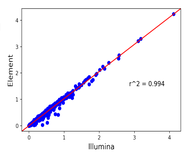
Total Clusters Identified *(various DESC resolutions)*

Resolution	Clusters in PBMC Control	Clusters in PBMC Depleted
0	0	0
0.1	5	7
0.4	8	9
0.8	10	12
1.2	13	16
1.6	16	17
2.0	19	20

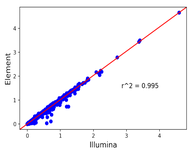
Single cell CRISPRclean depletion is a versatile method that can be easily inserted into many different experimental workflows, and can be sequenced on any platform:

Illumina vs Element

Comparison between a PBMC control and depleted library sequenced on Illumina versus Element.



The concordance between gene expression on the two sequencing platforms converges to 1 (log normalized average counts per gene).

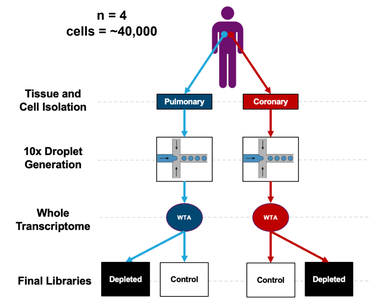


Applications

Assay	Experimentally Validated	Theoretically Applicable
Droplet based single cell RNA-seq (10x)	✓	
Plate based single cell RNA-seq (smartseq2, split-pooling)	✓	
Single cell Iso-seq and long read sequencing (SMRTbell)	✓	
Spatial Transcriptomics (10x Visium, seq-scope)	✓	
Spatial Transcriptomics (slide-seq, DBiT-seq, XYZeq)		✓
Single cell targeted capture (10x targeted gene expression)		✓
Single cell multi-omics (CITE-seq, SNARE-seq, scNMT-seq, sci-CAR, REAP-seq)		✓

Application

CRISPRclean identifies novel biological information when applied to single cell libraries of patient derived Vascular Smooth Muscle Cells (VSMC)



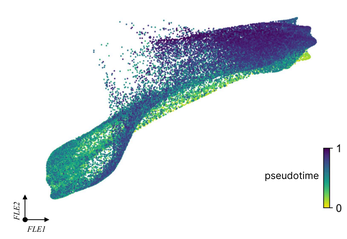
What are the transcriptional differences between Coronary and Pulmonary arteries?

Single cell libraries were generated from patient derived VSMCs. From each patient, VSMCs were isolated from both the coronary and pulmonary arteries. Depletion was performed on the WTA to ensure each control and depleted condition represented the same exact cells. The secondary analysis was performed on batch corrected data across 4 patients representing ~40,000 cells.

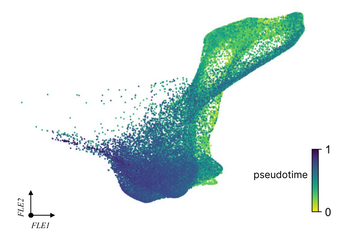
Depletion identifies an additional cell lineage with high confidence.

VSMCs are a plastic cell type that are known to fluctuate between many phenotypes. Performing trajectory analysis on these cells clearly identifies how the cells are transitioning from one state to another, taking multiple paths to get from initial cell state (yellow/green) to a terminal cell state (dark blue). Analysis was performed using CellRank[3] and scVelo[4] and all parameters were held constant between control and depleted.

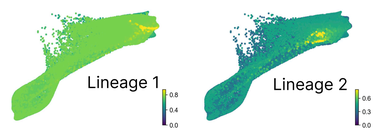
Control: *Pseudotime Trajectory*



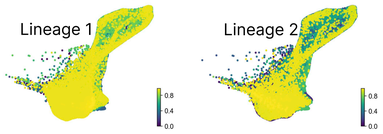
Depleted: *Pseudotime Trajectory*



Control: Identification of two lineages. Color scale represents the probability of a cell belonging to a specific lineage. >95% of the cells belong to only one lineage (lineage 1)

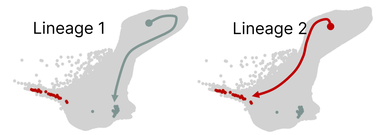


Depleted: Identification of two lineages. Color scale represents the probability of a cell belonging to a specific lineage. ~50% of cells belong to lineage 1. ~50% of cells belong to lineage 2.



Control: All of the VSMCs are transitioning transcriptionally in the same manner, terminating in the same phenotypic state.

Depleted: The VSMCs are transitioning transcriptionally in a distinct manner. Half are terminating in one phenotypic state, while the other half are terminating in a different state.

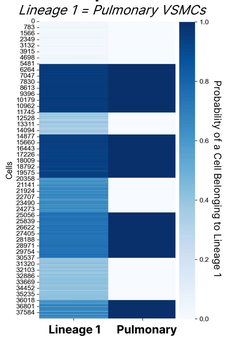


The two lineages identified in the depleted condition perfectly match whether the cells were from Coronary or Pulmonary arteries.

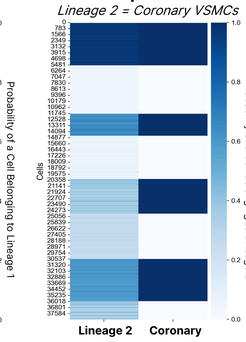
If a cell had a high probability of belonging to lineage 1, it was from a Pulmonary artery. If a cell had a high probability of belonging to lineage 2, it was from a coronary artery. (AUC=0.98)

Depletion identifies the difference between how Coronary VSMCs and Pulmonary VSMCs are changing transcriptionally. Even though the control condition represented the same exact cells, this distinction could not be made.

Depleted: *Lineage 1 = Pulmonary VSMCs*



Depleted: *Lineage 2 = Coronary VSMCs*



To learn more visit
jumpcodegenomics.com

Download the poster

Citations

1 – Aparicio, Luis, et al. "A random matrix theory approach to denoise single-cell data." *Patterns* 1.3 (2020): 100035
2 – Li, Xiangjie, et al. "Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis." *Nature communications* 11.1 (2020): 1-14.
3 – Lange, Marius, et al. "CellRank for directed single-cell fate mapping." *Nature methods* (2022): 1-12.
4 – Bergen, Volker, et al. "Generalizing RNA velocity to transient cell states through dynamical modeling." *Nature biotechnology* 38.12 (2020): 1408-1414.