JUMPCODE GENOMICS

TECH NOTE

Identify more genes per UMI with CRISPRclean® and Spatial Gene Expression

Boost informative transcripts and detect low expression genes in fresh frozen tissues

Introduction

Combining whole transcriptome gene expression profiles with histological information, providing spatially resolved gene expression, has sparked significant interest in spatial genomics. The ability to retain tissue organization and its cellular microenvironment while simultaneously probing the gene expression data has proved useful in various applications.[1-3] For example, identifying tumor-infiltrating lymphocytes (TILs) and mapping structural patterns have correlated to overall cancer patient survival.[1] However, challenges still exist in merging sequencing data with imaging readouts scalably. Sequencing read requirements are high and often dominated by ubiquitous transcripts that do not provide valuable biological information.

Jumpcode Genomics' CRISPRclean Single Cell RNA Boost Kit has proven utility in depleting common transcripts, such as ribosomal, mitochondrial, and non-variable gene content found in single-cell studies typically not useful in secondary analysis. The resulting increase in usable transcriptomic reads and detection of genes and UMIs enables researchers to boost useful information in their samples of interest. Here, we demonstrate the advantage of using CRISPRclean depletion with Visium Spatial Gene Expression to increase informative transcripts and detect low expression genes in two fresh frozen tissues: ovarian tumor and tonsil.

Protocol

Ovarian tumor and tonsil fresh frozen tissues were obtained from 10X Genomics and processed in collaboration. Briefly, the tissues were sectioned and fixed, then stained with hematoxylin and eosin stain (H&E) to allow visualization downstream. A permeabilization enzyme was then added to the tissue sections on a 10X Visium barcoded slide, releasing the poly-adenylated mRNA onto the primers on the spots located on the slide. A reverse transcription master mix was added to the permeabilized tissue sections and incubated to produce spatially barcoded, full-length cDNA from poly-A mRNA on the slide (see Visium Spatial Gene Expression User Guide for more detail).

The second strand mix was added, followed by denaturation and transfer of cDNA to a corresponding tube. The remaining steps in library construction were performed off chip and in-house at Jumpcode Genomics.

cDNA was amplified via PCR, followed by enzymatic fragmentation and size selection. End Repair, A-tailing, Adapter Ligation and cleanup were all performed. cDNA from the same tissue was split into two separate libraries: One library was prepared using the Visium Spatial Gene Expression library preparation, and a separate library was prepared with the Visium Spatial Gene Expression library preparation and CRISPRclean Single Cell RNA Boost (Figure 1).

Highlight

- 46-50% more genes per UMI were identified using CRISPRclean® Single Cell RNA Boost in conjunction with Visium® Spatial Gene Expression for tonsil and ovarian tumor fresh frozen tissue sections.

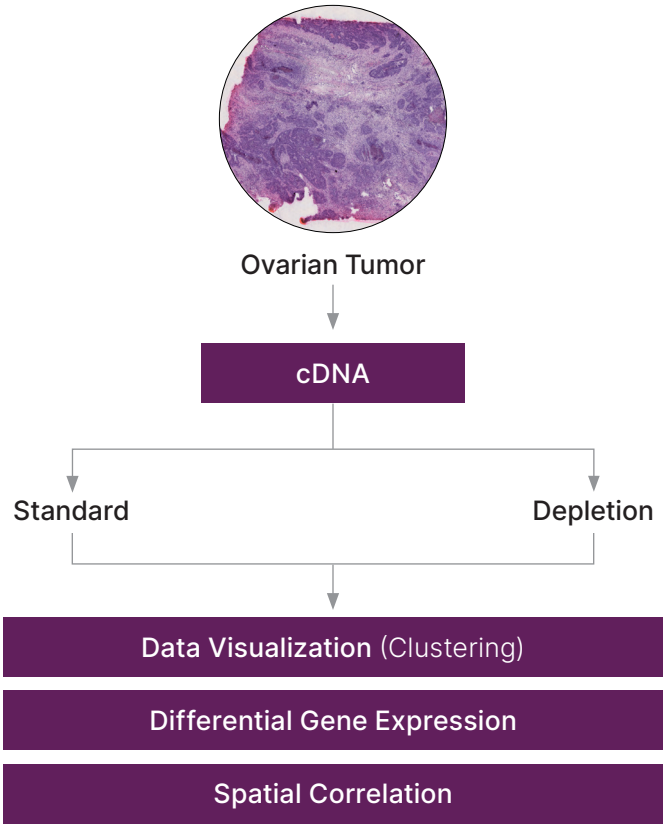

Figure 1: cDNA splits were used to compare the standard Visium Spatial Gene Expression results to the depleted condition in which samples were processed with Visium Spatial Gene Expression plus CRISPRclean Single Cell RNA Boost. Both samples were sequenced to 80,000 reads per spot and processed with the 10x Space Ranger platform, including clustering, differential gene expression, and spatial correlation analysis.

Four regions for removal were targeted: ribosomal and mitochondrial genes in the transcript, non-variable genes in the transcript, ribosomal RNA in the genomic reads, and unaligned reads in genomic sequences. This content is typically uninformative for secondary analysis and often removed informatically. By removing this content prior to sequencing, reads can be re-allocated to informative transcripts, boosting coverage and reducing sequencing costs. Content targeted for removal by CRISPRclean Single Cell RNA Boost is shown in Figure 2.

Results

Depletion with CRISPRclean Single Cell RNA Boost enhances per spot resolution, enabling detection of more genes.

Samples processed with CRISPRclean depletion recovered more information per spot, as quantified in Figure 3A&B. In Figure 3A, the x-axis shows the counts of genes divided by counts of UMI. In the depleted condition, we recover more UMI counts per gene, as evidenced by the shift in the curve to the right of the plot. At 1,000 reads per spot, for example, we would recover 320 genes in the control and 480 genes in the depleted condition.

Similarly, in Figure 3B, the ovarian tumor sample results show the same phenomenon. The depleted sample (orange) population is shifted to the right, showing more UMIs per gene in the depleted condition compared to the control condition. At 1,000 reads per spot, the control recovers 370 genes, while the depleted condition

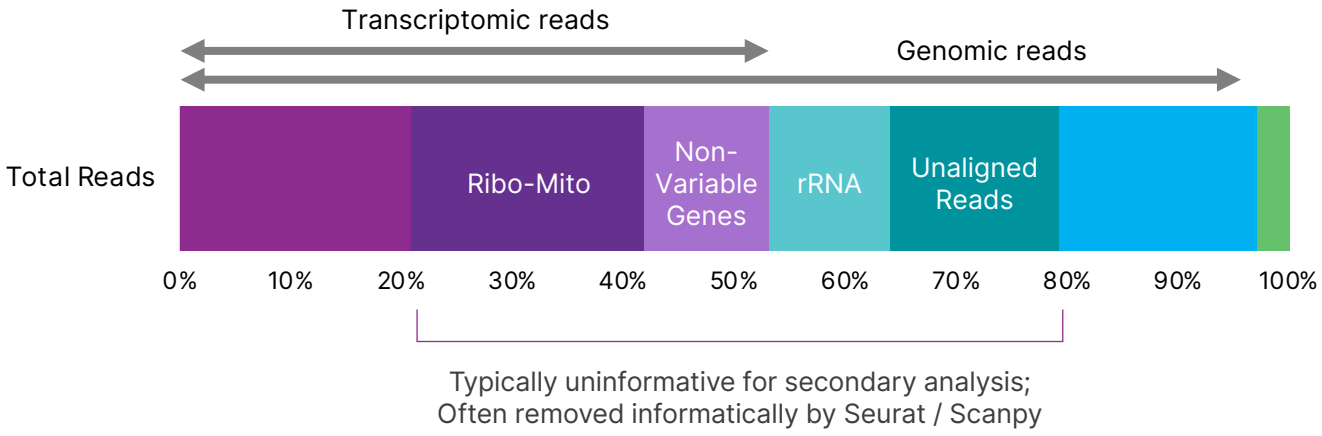

Figure 2: The total reads, by content, of a typical single-cell sequencing run are shown with the distribution of reads plotted on the x-axis. Four regions for removal are highlighted. Ribosomal and mitochondrial (Ribo-Mito) genes and non-variable genes from the transcriptomic reads are the first two regions highlighted. Ribosomal RNA (rRNA) and unaligned reads from genomic reads are also targeted for removal.

recovers 540 genes. This coverage increase results from removing sequencing reads from uninformative transcripts, such as ribosomal, mitochondrial, and non-variable gene content, and remapping those reads onto informative transcripts.

To show that the depleted condition will continue to give more genes per UMI at all sequencing depths, we plotted the number of reads for both tissue types in Figure 4 on the x-axis and compared that to the sequencing saturation on the y-axis. The depleted condition shows more saturation or complexity of the library sequence at all sequencing depths compared to the control. Further, to achieve a specific sequencing

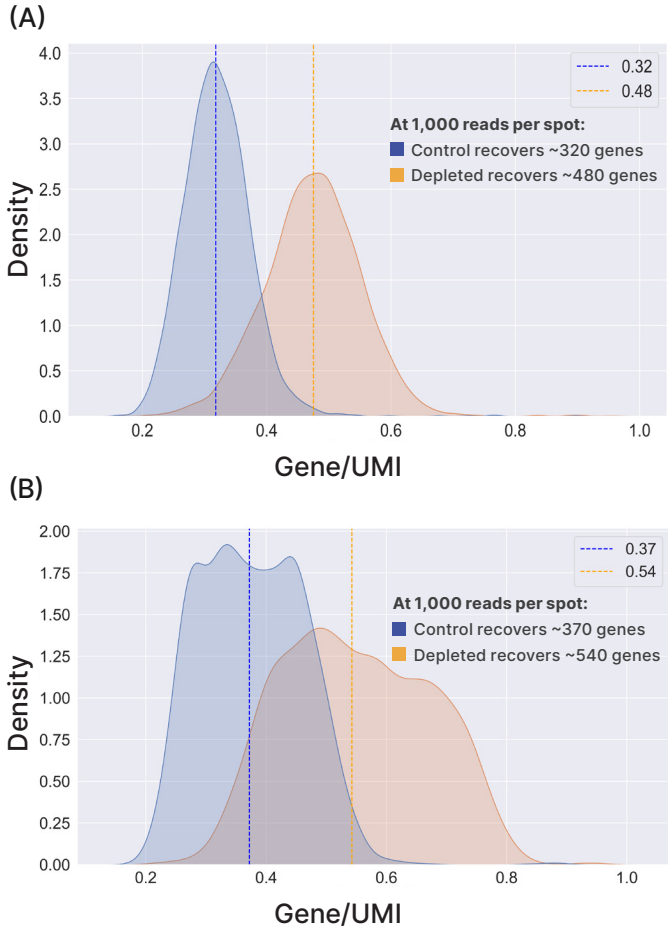

Figure 3: The genes per UMI (x-axis) were plotted against the tonsil tissue sample's kernel density estimate (y-axis). **(A)** At 1,000 reads per spot, 50% more genes were detected in the depleted tonsil sample (orange) than in the control (blue). **(B)** The genes per UMI (x-axis) were plotted against the ovarian tumor tissue sample's kernel density estimate (y-axis). At 1,000 reads per spot, 46% more genes were detected in the depleted ovarian tumor sample (orange) relative to the control (blue).

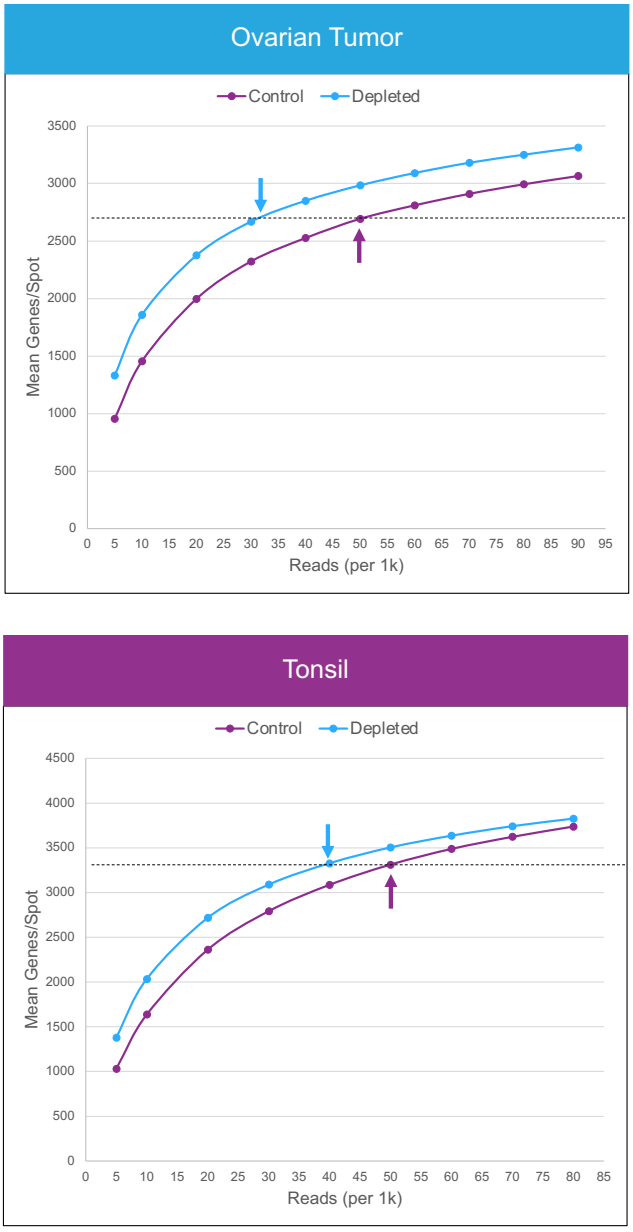

Figure 4: Reads for both ovarian tumor and tonsil tissue are plotted on the x-axis and compared to sequencing saturation on the y-axis. The depleted condition (blue) shows higher sequencing saturation for every increase in read depth compared to the control condition (purple). The arrows indicate that for the recommended 50,000 reads/spot in the control condition, the same information can be found in the depleted condition by sequencing half the number of reads or 25,000 reads/spot.

saturation, the researcher only needs half the number of reads for the depleted condition compared to the control condition. Thus, the researcher can choose to save sequencing costs by sequencing half of the recommended sequencing depth to obtain the same information or sequence at the same depth and reveal more information (Figure 4).

Depletion identifies two additional clusters with biologically relevant lowly expressed markers.

Additionally, two additional clusters were identified when comparing the results of semi-supervised clustering using Seurat v4 on ovarian tumor samples (Figure 5A). The control sample, on the left, has two clusters that split when analyzing the depleted samples. Cluster 4, in the control, is divided into clusters 6 and 9 in the depleted sample, and cluster 2 in the control is split into clusters 3 and 8 in the depleted sample. Because of the removal of ribosomal, mitochondrial, and non-variable gene content in the depleted sample, the increased coverage of transcriptomic data in the depleted sample allows the identification of more unique reads, resulting in the detection of two additional clusters of cells.

Further, looking at the gene expression data, the top spatially enriched genes for the control show similarities but also some key differences compared to the depleted condition. The sepal scores give a number to quantify gene colocalization with spatial structure in tissue. A high sepal score corresponds to a gene with a high correlation to a spatial pattern.[5] In the depleted condition, the gene S100A7 shows a high sepal score of 6.178. However, that gene is not in the top spatially enriched genes list in the control sample. Exploring this further, we confirmed that the S100 family of genes comprises proteins involved in cellular processes implicated in ovarian cancer6. Most S100 family genes were overexpressed in ovarian cancer compared to normal tissues. As seen in Figure 5B, both

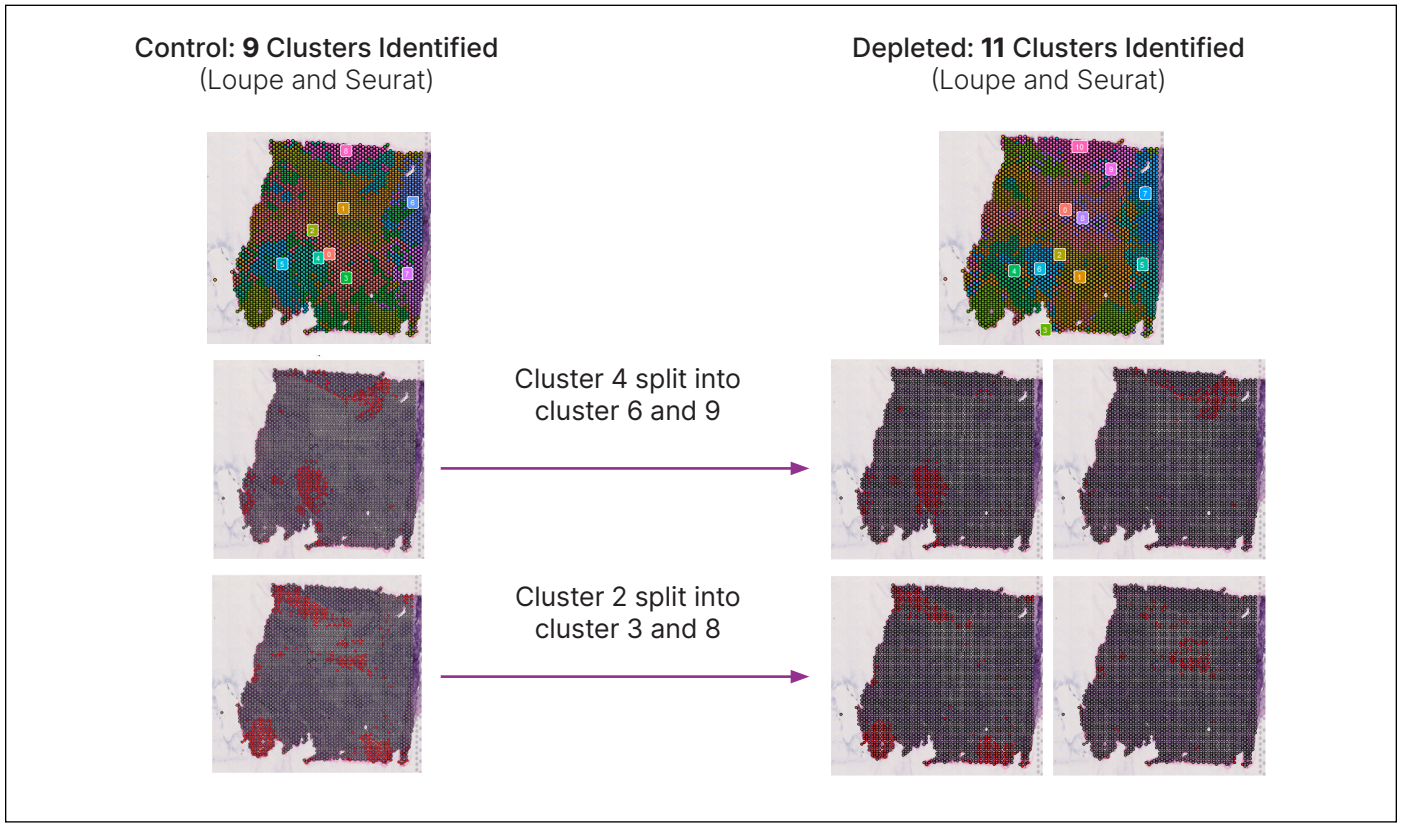

Figure 5A: Semi-supervised clustering was performed using Seurat v4. Clustering performance with standard 10x Visium generated libraries (left) and with CRISPRclean depletion (right). Two additional spatially resolved clusters were identified with CRISPRclean depletion for Ovarian Tumor samples.

S100A7 and S100A9 are found in the depleted condition for the ovarian cancer tissue samples, while the control condition does not include them in the top spatially enriched genes list. This demonstrates the power of depletion in bringing important genetic expression data to the surface: data is often overshadowed in non-depleted data due to over-sequencing of ribosomal, mitochondrial, and non-variable genetic content.

Conclusion

Spatial genomics promises to shine a light on transcriptomic information directly in the tissue microenvironment. Often, the most interesting information lies in the change of lower abundance transcripts. Jumpcode Genomics' CRISPRclean Single Cell RNA Boost kit improves detection of these transcripts by effectively removing ubiquitous and uninformative sequences.

Control	
	sepal score
TMEM52	2.539
XDH	2.515
ZNF670	2.514
KCNN3	2.123
HIST2H2AC	2.060
TENT5C	2.044

Depleted	
	sepal score
S100A7	6.178
TMEM52	2.747
NBPF14	2.251
AL365361.1	2.135
UBXN10-AS1	1.911
S1000A9	1.871

Figure 5B: The top spatially enriched genes in the depleted sample are relevant to ovarian tumors. These genes were not found in the control condition.

To learn more, visit jumpcodegenomics.com

Ordering information

Catalog	Product name	Samples
KIT1018	CRISPRclean Single Cell RNA Boost Kit	24

References

1. Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Samaras D, Shroyer KR, Zhao T, Batiste R, Van Arnam J; Cancer Genome Atlas Research Network, Shmulevich I, Rao AUK, Lazar AJ, Sharma A, Thorsson V. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Rep. 2018 Apr 3;23(1):181-193.e7. doi: 10.1016/j.celrep.2018.03.086. PMID: 29617659; PMCID: PMC5943714.
2. Grauel AL, Nguyen B, Ruddy D, Laszewski T, Schwartz S, Chang J, Chen J, Piquet M, Pelletier M, Yan Z, Kirkpatrick ND, Wu J, deWeck A, Riester M, Hims M, Geyer FC, Wagner J, MacIsaac K, Deeds J, Diwanji R, Jayaraman P, Yu Y, Simmons Q, Weng S, Raza A, Minie B, Dostalek M, Chikkegowda P, Ruda V, Iartchouk O, Chen N, Thierry R, Zhou J, Pruteanu-Malinici I, Fabre C, Engelman JA, Dranoff G, Cremasco V. TGFβ-blockade uncovers stromal plasticity in tumors by revealing the existence of a subset of interferon-licensed fibroblasts. Nat Commun. 2020 Dec 9;11(1):6315. doi: 10.1038/s41467-020-19920-5. PMID: 33298926; PMCID: PMC7725805.
3. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, Tarish F, Tanoglidi A, Vickovic S, Larsson L, Salmén F, Ogris C, Wallenborg K, Lagergren J, Ståhl P, Sonnhammer E, Helleday T, Lundeberg J. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat Commun. 2018 Jun 20;9(1):2419. doi: 10.1038/s41467-018-04724-5. PMID: 29925878; PMCID: PMC6010471.
4. Bai Y, Li LD, Li J, Lu X. Prognostic values of S100 family members in ovarian cancer patients. BMC Cancer. 2018 Dec 17;18(1):1256. doi: 10.1186/s12885-018-5170-3. PMID: 30558666; PMCID: PMC6296138.
5. Anderson A, Lundeberg J. sepal: Identifying Transcript Profiles with Spatial Patterns by Diffusion-based Modeling. Bioinformatics. 2021 Mar 11;37(17):2644–50. doi: 10.1093/bioinformatics/btab164. Epub ahead of print. PMID: 33704427; PMCID: PMC8428601.
6. Lin M, Xia B, Qin L, Chen H, Lou G. S100A7 Regulates Ovarian Cancer Cell Metastasis and Chemoresistance Through MAPK Signaling and Is Targeted by miR-330-5p. DNA Cell Biol. 2018 May;37(5):491-500. doi: 10.1089/dna.2017.3953. Epub 2018 Feb 27. PMID: 29485916.