TECH NOTE

Increase scRNA-seq coverage of low-expression genes and isoforms with depletion

Combining PacBio® MAS-Seq and Jumpcode Single Cell RNA Boost Kit improves isoform detection sensitivity

Highlights

- 2.6-fold boost in transcriptomic reads
- Higher numbers of novel isoforms discovered – 30,655 in depleted compared to 27,732 control
- 1.6-fold more highly variable genes detected
- Identification of mast cells, a rare cell type in PBMCs, through identification of genes such as CPA3, a canonical mast cell marker

Introduction

Single-cell RNA sequencing (scRNA-seq) technologies are revolutionizing transcriptome studies. Their unique ability to characterize the transcriptome of a cell on an individual level provides unprecedented detail.[1,2] Using PacBio long-read sequencing, alternative splicing events within cells can be discovered at a higher rate than short-read sequencing.[3] To scale up this method and increase user adoption, however, the cost of sequencing and the ability to detect low-expression genes of interest need to be improved. By combining the Jumpcode Single Cell RNA Boost Kit—a CRISPR-Cas9 system that removes uninformative genes from scRNA-seq libraries before sequencing (Figure 1) with long-read single-cell RNA sequencing (MAS-Seq)—we demonstrate how the removal of uninformative genes greatly increases the discovery power of isoforms – and rare cell types – for single-cell transcriptomics.

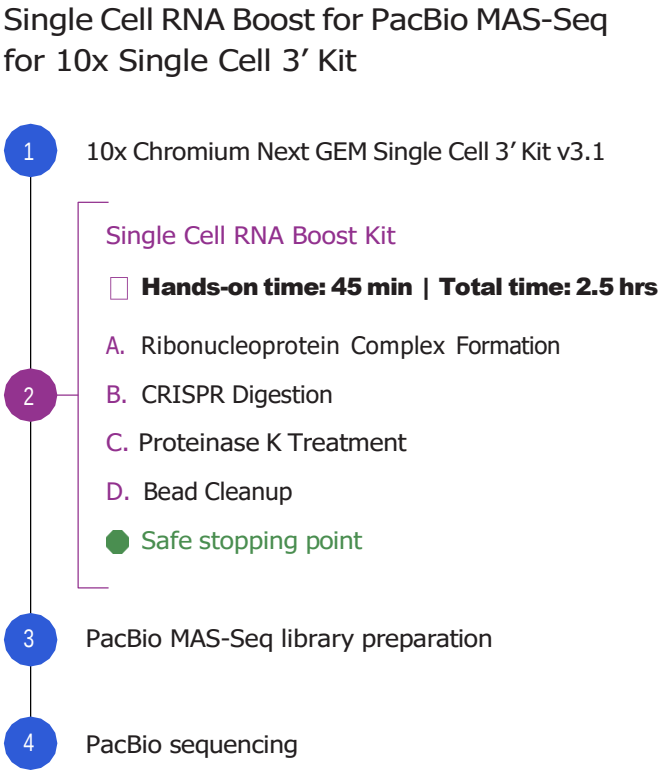

Figure 1: Overview of PacBio MAS-Seq + Jumpcode depletion technology. cDNA was generated using the 10x Chromium Next GEM Single Cell 3' Kit 3.1 workflow, followed by processing samples with the Jumpcode Single Cell RNA Boost Kit. The depleted cDNA was then made into libraries using the PacBio *MAS-Seq for 10x Single Cell 3'* kit and sequenced on PacBio systems.

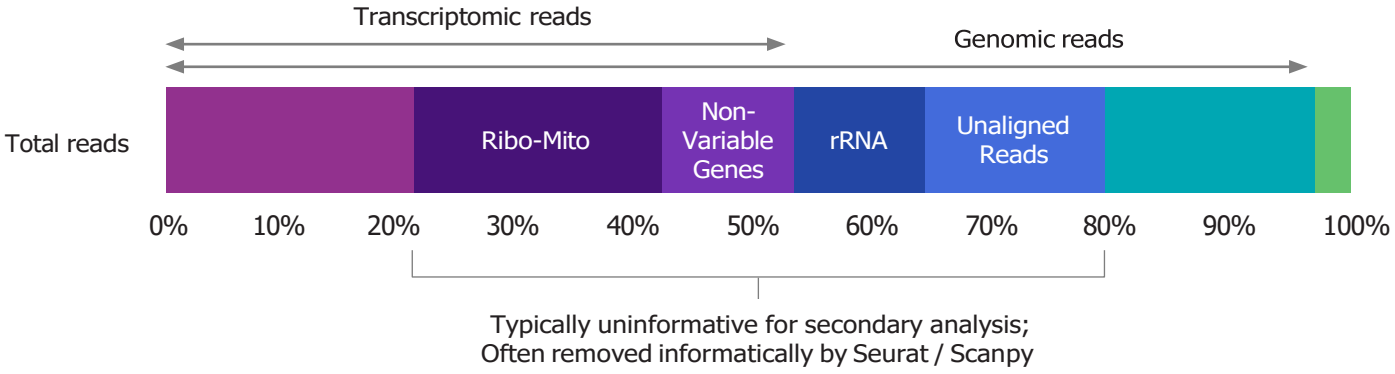

Figure 2: Visual representation of targeted content for depletion in Single Cell RNA Boost Kit. Results represent typical read distribution from a single-cell sequencing run. Ribo/Mito and Non-Variable Genes in the transcriptome are targeted for removal along with unaligned reads and rRNA in the genomic content. Full depletion content is available at revvity.com.

Depleting uninformative genes in 10x single cell cDNA for long-read isoform sequencing

The CRISPR guide RNAs in the Single Cell RNA Boost Kit are designed to deplete uninformative genes commonly found in 10x single cell data. By analyzing a cohort of publicly available data from various sources, roughly 30-50% of reads were found to align to the genome but not the transcriptome and, thus, are conventionally ignored by analysis tools. By tailoring guides to deplete these genomic intervals, along with the highest-expressed protein-coding ribosomal and mitochondrial genes, we can redistribute ~50% of reads to informative transcripts (Figure 2).

RNA was extracted from 12,000 PBMCs and processed using the Chromium Next GEM Single Cell 3' v3.1 dual index kit from 10x Genomics. Samples were then processed with the Single Cell RNA Boost Kit from Jumpcode Genomics, keeping a set of non-depleted samples aside to serve as a non-depleted control condition. The protocols used for this experiment can be found on the Jumpcode and PacBio websites. Single cell MAS-Seq libraries were then prepared for control and depleted conditions. The resulting libraries were sequenced on the PacBio Sequel® II sequencing platform using one SMRT® Cell 8M per sample. The resulting data from the control and depleted samples was analyzed with PacBio SMRT® Link software (v11.1) and Seurat[4] for secondary analysis. IGV was used to visualize the coverage of specific genes.

Jumpcode-depleted single cell MAS-Seq data increased usable reads and boosted UMIs and genes per cell

The Single Cell RNA Boost Kit successfully removed 86% of the targeted genes (Figure 3), decreasing the percentage of aligned reads attributed to the target genes from 56% to 8%. This represents a 2.6-fold increase in "usable reads" compared to MAS-Seq data

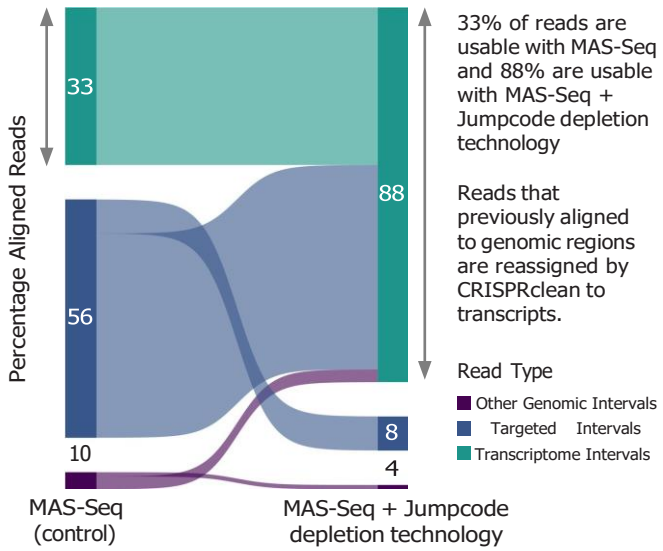

Figure 3: Jumpcode depletion increased usable reads in single cell MAS-Seq data. The percent of reads mapping to targeted genes and intervals was calculated for control and depleted samples. Reads from targeted regions (blue) are greatly reduced by the Jumpcode depletion method resulting in an increase in reads mapping to transcriptome intervals (green).

run without Jumpcode depletion—boosting the signal to noise in the sample and allowing increased opportunity for sequencing of low-expression genes and detection of novel isoforms. In the depleted sample, 88% of reads were mapped to transcriptome intervals compared to only 33% in the control sample.

By removing highly abundant content, reads are reassigned to lower expression targets resulting in a boost in UMIs and genes per cell. Compared to the control sample, the depleted sample had a 1.5-fold boost in the number of UMIs per cell and genes per cell (Figure 4).

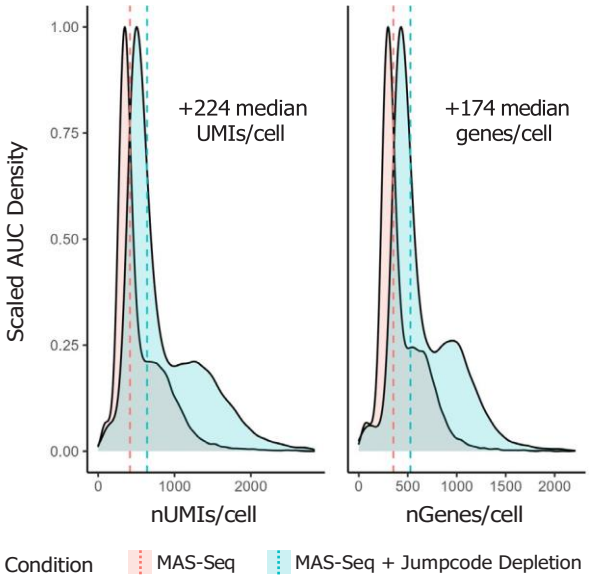

Figure 4: Boost in UMIs and genes per cell captured with Jumpcode depletion. The MAS-Seq control sample (pink) compared to the MAS-Seq + Jumpcode depleted sample (turquoise) showed a 1.5-fold increase in the number of UMIs and genes per cell.

We compared the control and depleted libraries for total and different types of isoforms discovered using the SQANTI transcript classification system implemented in SMRT Link v11.1. A full splice match (FSM) is defined as a transcript that matches a reference transcript at all splice junctions, while an incomplete splice match (ISM) is defined as a transcript that matches consecutive, but not all, splice junctions of the reference transcript. Novel in catalog (NIC) transcripts contain new combinations of already annotated splice junctions or novel splice junctions formed from the already annotated donor and acceptor sites. Novel Not in Catalogue (NNC) transcripts use at least one donor and/or acceptor site.[6]

The number of isoforms per cell increased by 1.5X in the depleted condition compared to the control, as seen in Figure 6, highlighting the difference between pseudo-bulk and single-cell analysis.

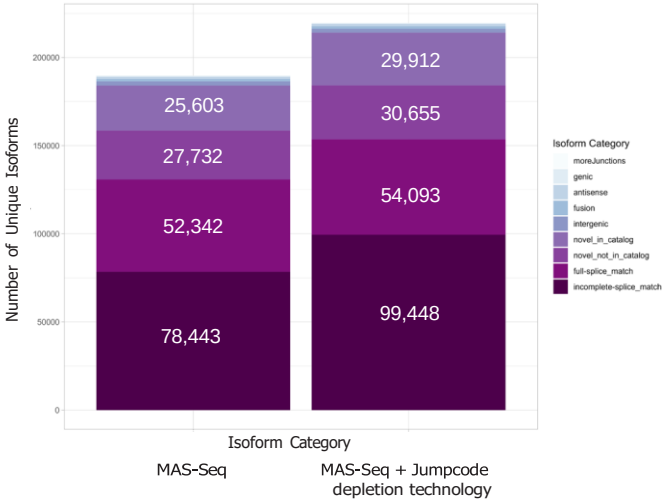

Figure 5: More isoforms are identified in Jumpcode depletion-treated libraries than in control samples. Unique isoforms were tallied and classified using the SQANTI transcript classification nomenclature.

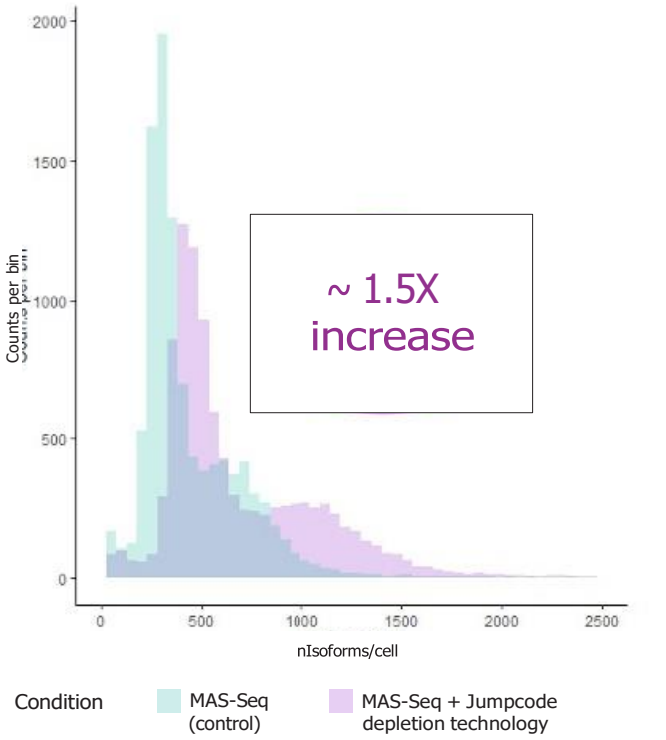

Figure 6: 1.5X boost in isoforms/cell detected. Using MAS-Seq + Jumpcode depletion results in a 1.5X boost in the number of isoforms per cell detected when compared to using MAS-Seq alone on PBMC libraries.

Jumpcode-depleted single cell MAS-Seq data identified rare cell types

We compared the number of highly variable genes (HVGs) between the control and depleted samples. HVGs exhibit high cell-to-cell variation, meaning they are highly expressed in some cells and lowly in others. By focusing on these genes in downstream analysis, differences from cell to cell can be highlighted. In the control sample, 1279 HVGs were found compared to 2067 HVGs in the depleted sample. This boost in the number of HVGs resulted in four additional clusters in the UMAP plots for the depleted samples (Figure 7). The increased resolution of biological data in the depleted sample shows the power of the Jumpcode depletion method: it allows additional biological insights into samples by removing the noise, resulting in a 1.6X increase in the number of HVGs found in the depleted sample compared to the control.

We show an example of the CPA3 gene between short-read (Illumina®) sequencing, MAS-Seq alone, and MAS-Seq + Jumpcode depletion conditions (Figure 8). CPA3 is a known canonical gene for identifying rare cell types such as mast cells.[9] Short reads failed to capture this biologically relevant gene. Both long-read samples captured the CPA3 gene. However, the Jumpcode depletion sample has coverage on all exons, allowing identification of the additional cell cluster in the UMAP plots.

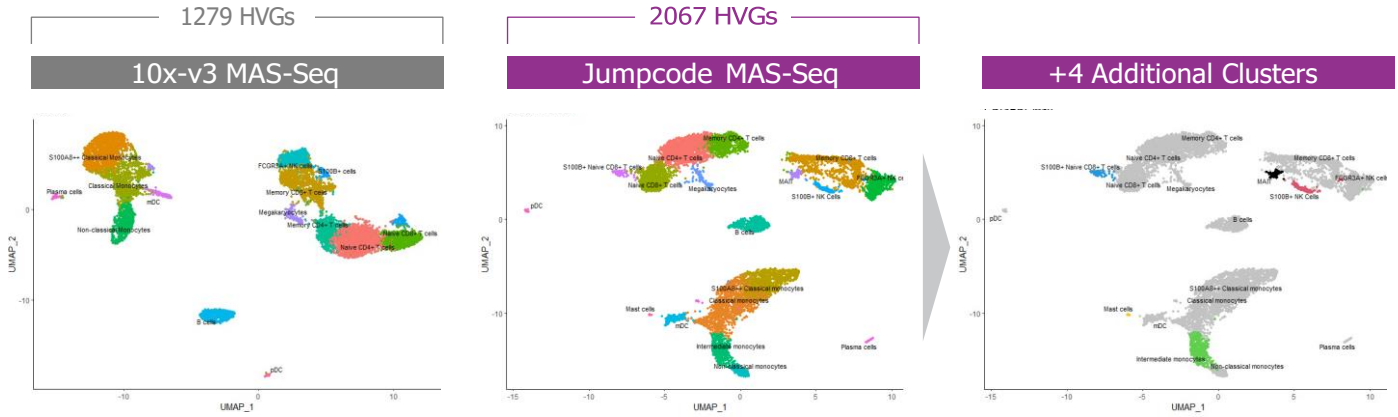

Figure 7: Jumpcode-depleted MAS-Seq data identifies four additional rare cell clusters in PBMC. Depleted samples have 4 additional clusters under the same analytical conditions as the control because of the increased detection of highly variable genes (HVGs).

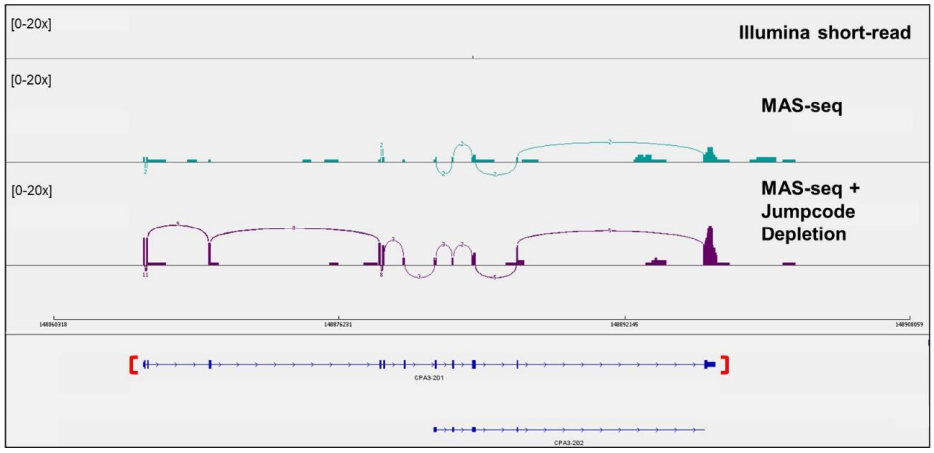

Figure 8: Sashimi plot for the CPA3 gene, a mast cell marker, has coverage of all exons in the MAS-Seq + Jumpcode depletion condition. Rare and difficult-to-sequence genes obtain no coverage using Illumina short-read sequencing, low coverage using MAS-Seq alone, but full coverage of all exons using MAS-Seq + Jumpcode depletion.

Conclusions

Single-cell transcriptome studies, while offering unprecedented characterization of transcript expression at the single cell level, suffer from decreased sensitivity since data from highly-expressed genes can obscure signals from lower-expressed genes. By removing highly-expressed genes before sequencing, the Jumpcode depletion method preserves sequencing space for informative molecules. The results show a 2.6X boost in transcriptomic reads with higher number of novel isoforms depleted compared to the control. With 1.6X more highly variable genes detected, four additional clusters were identified—including mast cells.

To learn more, visit revvity.com

Ordering information

Catalog	Product name	Samples
NOVA-512850	Single Cell RNA Boost Kit	24

References

1. Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, Fronick CC, Fulton RS, Church DM, Ley TJ. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat Commun. 2019 Aug 14;10(1):3660.
2. Moreno-Villanueva M, Zhang Y, Feiveson A, Mistretta B, Pan Y, Chatterjee S, Wu W, Clanton R, Nelman-Gonzalez M, Krieger S, Gunaratne P, Crucian B, Wu H. Single-Cell RNA-Sequencing Identifies Activation of TP53 and STAT1 Pathways in Human T Lymphocyte Subpopulations in Response to Ex Vivo Radiation Exposure. Int J Mol Sci. 2019 May 10;20(9):2316.
3. Al'Khafaji AM, Smith JT, Garimella KV, Babadi M, Sade-Feldman M, Gatzen M, et al. High-throughput RNA isoform sequencing using programmable cDNA concatenation [Internet]. bioRxiv. 2021:2021.10.01.462818 Available from: https://www.biorxiv.org/content/10.1101/2021.10.01.462818v1.
4. "Seurat - Guided Clustering Tutorial," January 11, 2022, https://satijalab.org/seurat/articles/pbmc3k_tutorial.html.
5. Huang KK, Huang J, Wu JKL, Lee M, Tay ST, Kumar V, Ramnarayanan K, Padmanabhan N, Xu C, Tan ALK, Chan C, Kappei D, Göke J, Tan P. Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. Genome Biol. 2021 Jan 22;22(1):44.
6. Veiga DFT, Nesta A, Zhao Y, Mays AD, Huynh R, Rossi R, Wu TC, Palucka K, Anczukow O, Beck CR, Banchereau J. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv. 2022 Jan 21;8(3).
7. Palmer CR, Liu CS, Romanow WJ, Lee MH, Chun J. Altered cell and RNA isoform diversity in aging Down syndrome brains. Proc Natl Acad Sci U S A. 2021 Nov 23;118(47).
8. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, Edelmann M, Ezkurdia I, Vazquez J, Tress M, Mortazavi A, Martens L, Rodriguez-Navarro S, Moreno-Manzano V, Conesa A. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018 Feb 9;28(3):396–411. doi: 10.1101/gr.222976.117. Epub ahead of print. Erratum in: Genome Res. 2018 Jul;28(7):1096. PMID: 29440222; PMCID: PMC5848618.
9. Jiang, J., Faiz, A., Berg, M., Carpaij, O. A., Vermeulen, C. J., Brouwer, S., Hesse, L., Teichmann, S. A., Ten Hacken, N. H. T., Timens, W., van den Berge, M., & Nawijn, M. C. (2020). Gene signatures from scRNA-seq accurately quantify mast cells in biopsies in asthma. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology, 50(12), 1428–1431. https://doi.org/10.1111/cea.13732
10. gr.222976.117. Epub ahead of print. Erratum in: Genome Res. 2018 Jul;28(7):1096. PMID: 29440222; PMCID: PMC5848618.