

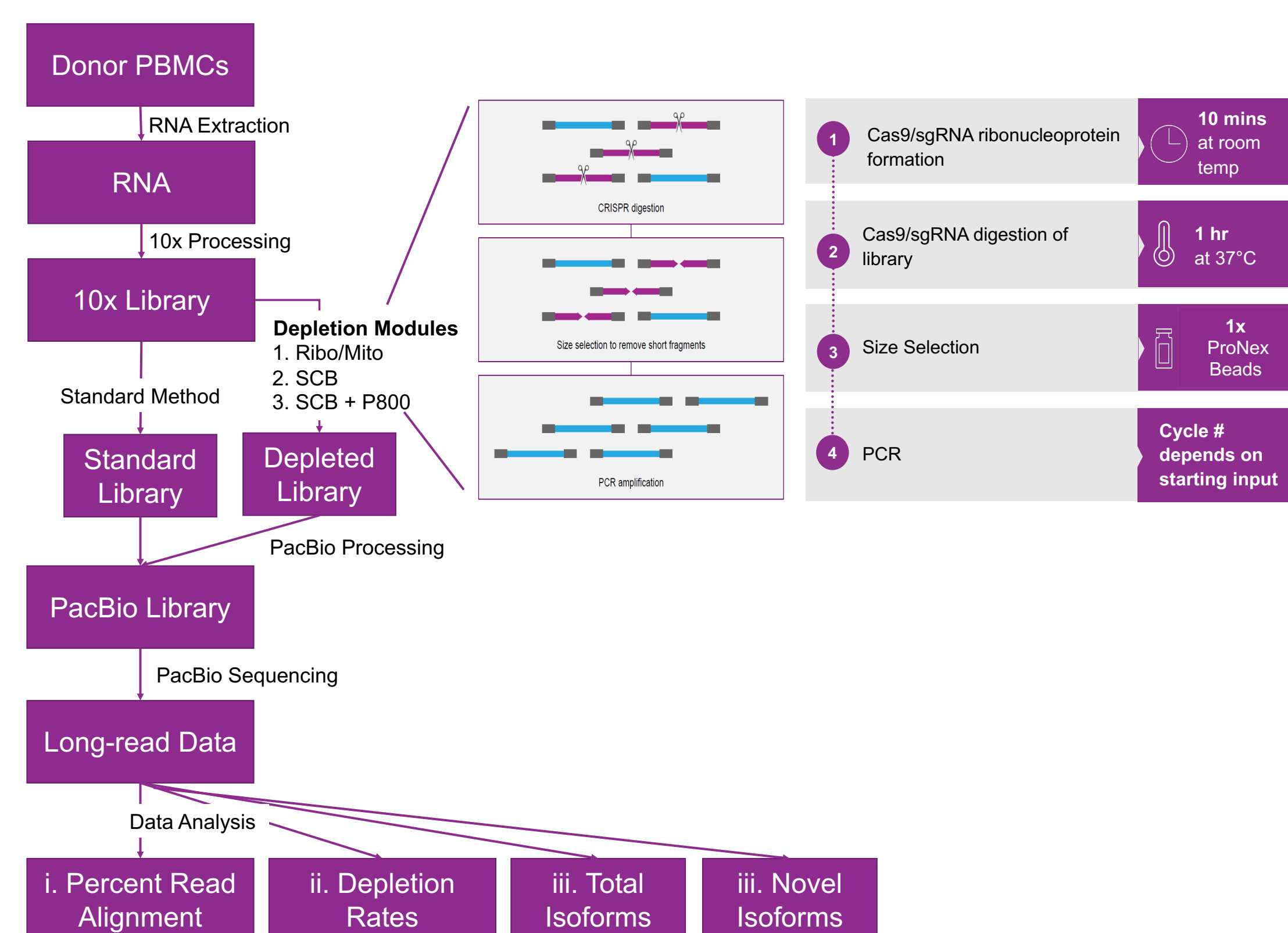
# Improving Detection Sensitivity for PacBio ISO-Seq using CRISPRclean®

Jon R. Armstrong, Jeffrey Deason, Austin Crinklaw, Sridhar Ranganathan, Azeem Siddique, Keith Brown.  
Jumpcode Genomics, San Diego CA



## Introduction

Single cell RNA sequencing (scRNA-seq) technologies have helped revolutionize transcriptome studies with the ability to characterize the transcriptome of a cell in unprecedented detail<sup>1,2</sup>. Of particular interest is the exploration of alternative splicing in single cells<sup>1-3</sup> and both short and long read systems have been utilized. However, existing sequencing platforms do not scale creating limitations in isoform quantitation and discovery. It is challenging for short read approaches to span successive exon boundaries and long read systems do not have sufficient output to characterize lower expressed isoforms. Even if one were to overcome these limitations, transcripts from highly expressed genes overwhelm biological signals from alternative splicing events. Here, we present an approach based on the CRISPR-Cas9 system, marketed as CRISPRclean that removes uninformative fragments from scRNA-seq libraries prior to sequencing (Figure 1), greatly increasing informative reads in scRNA-seq data and enabling discovery of additional isoforms and cell types.



**Figure 1. Overview of CRISPRclean technology.** Double stranded full length cDNA products are cleaved using CAS9 and specifically designed guide RNAs.

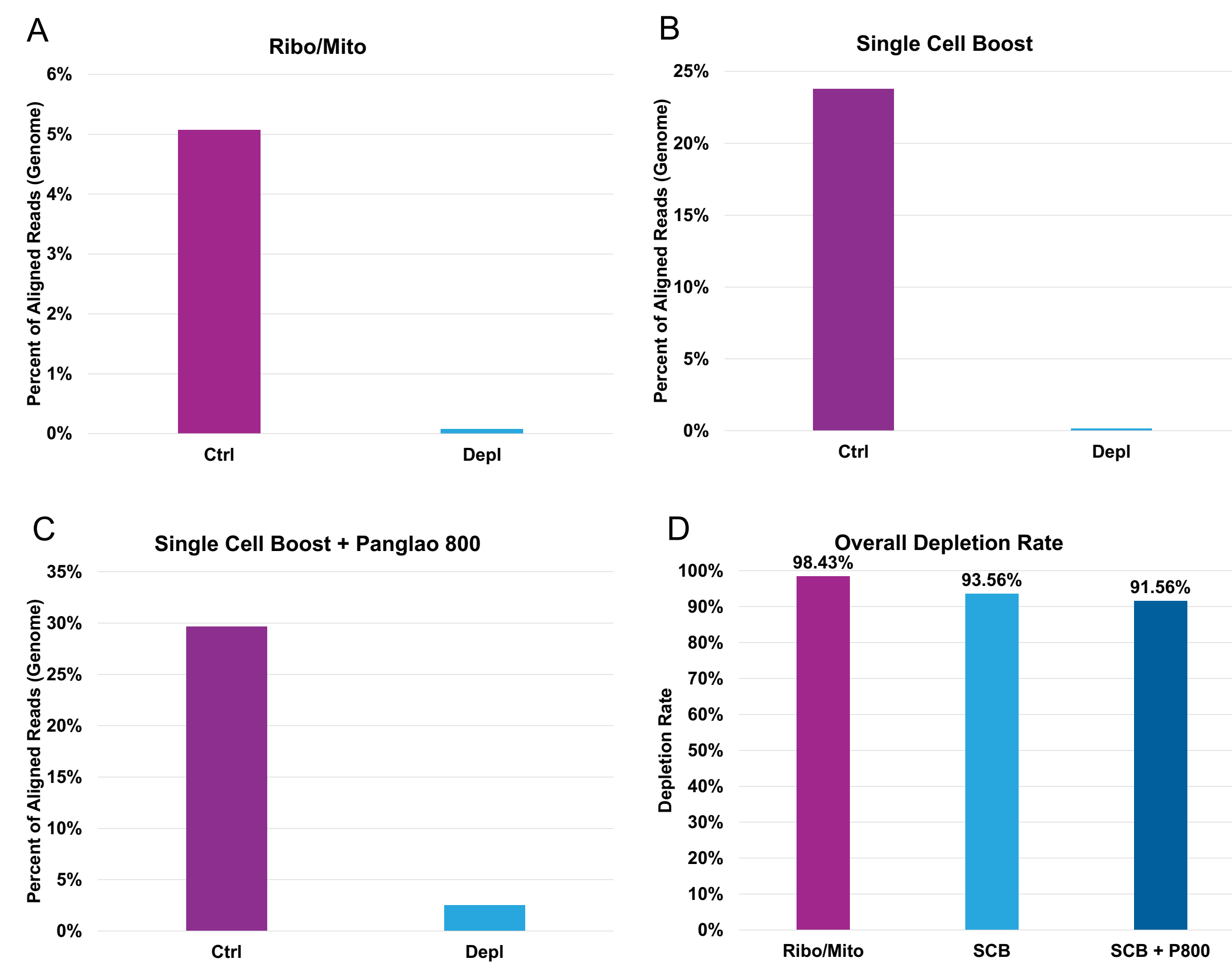
## Methods

RNA was extracted from donor PBMC samples and processed through the 10x single cell system using the Chromium Next GEM Single Cell 3' v3.1 dual index kit. Single-Cell Iso-Seq libraries were prepared using the PacBio SMRTbell Express Template Prep Kit 2.0 with and without depletion. CRISPR guide RNAs were designed against approximately 1699 genes and genomic intervals identified for removal from sequencing libraries. The guides were tested for depletion in three modules consisting of guides targeting coding ribosomal and mitochondrial genes (Ribo/Mito), single cell boost targeting unaligned reads, ribosomal, mitochondrial, non-variable genes (SCB) and SCB + 800 highest ubiquity genes from the PanglaoDB (SCB + P800). Approximately 130,000 unique guide RNAs were synthesized for testing in the CRISPRclean method. The resulting libraries were sequenced on the PacBio Sequel II Sequencing platform using 1 SMRT Cell 8 M per sample. The resulting data from control (CTRL) and depleted (DEPL) samples was analysed for i. percentage of reads aligning to each targeted region, ii. depletion rates of targeted molecules, iii. total isoforms, and iv. novel isoforms. (Figure 1).

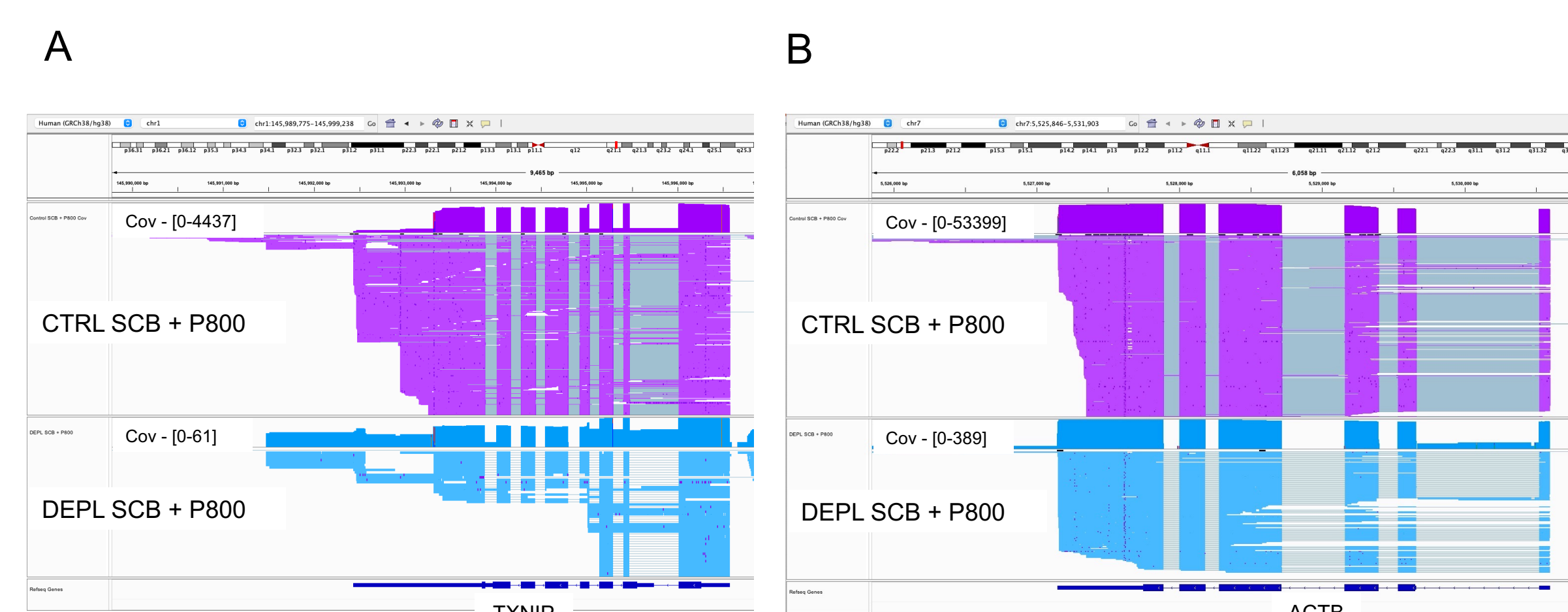
## Results

### CRISPRclean efficiently removes targeted molecules

Libraries were treated with and without CRISPRclean plus depletion. Multiple guide sets were tested independently and in combination. Figure 2 shows that each guide set used in the CRISPRclean workflow efficiently removes targeted molecules from single cell Iso-seq sequencing libraries. IGV coverage visualization of two genes, Thioredoxin Interacting Protein (TXNIP) and Actin Beta (ACTB), included in the single cell boost and Panglao 800 sets are shown in Figure 3 and exhibit a reduction in read coverage of approximately 100-fold, confirming the efficacy of the CRISPRclean method.



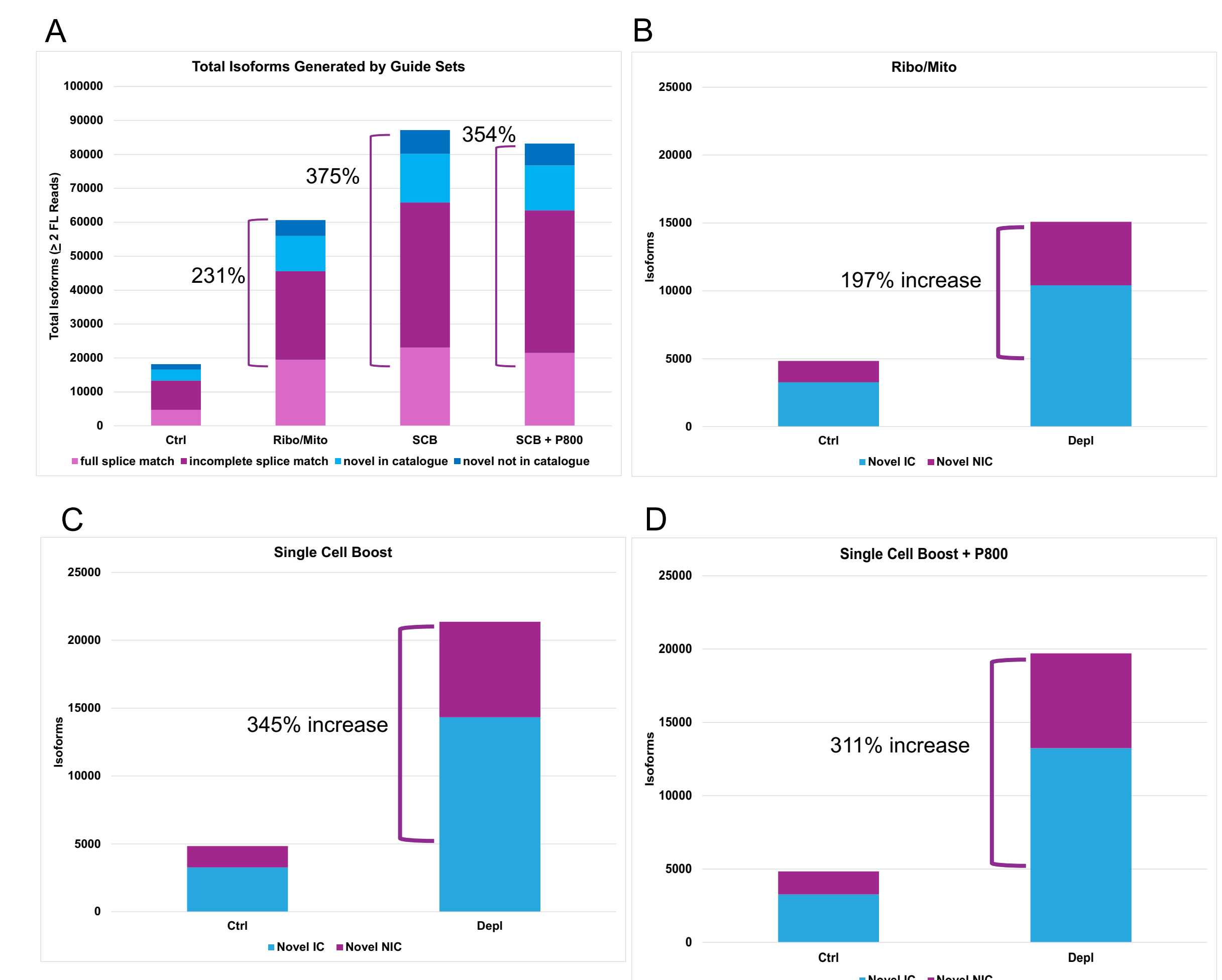
**Figure 2. Depletion with guide modules shows reduced read alignments and high depletion rates.** The percent of reads mapping to targeted genes and intervals (y-axis) was computed for control (CTRL, purple) and depleted (DEPL, blue) samples. Reads from targeted regions are almost completely removed by the CRISPRclean method. Multiple guide set combinations were used for depletion, A. coding ribosomal and mitochondrial genes - Ribo/Mito; B. single cell boost genes; C. single cell boost and Panglao 800 - P800 genes. Figure D shows the depletion rate (y-axis) is greater than 90% for each module (coding ribosomal and mitochondrial genes - Ribo/Mito, purple; single cell boost genes - SCB, light blue; single cell boost and Panglao 800 genes - SCB + P800, dark blue).



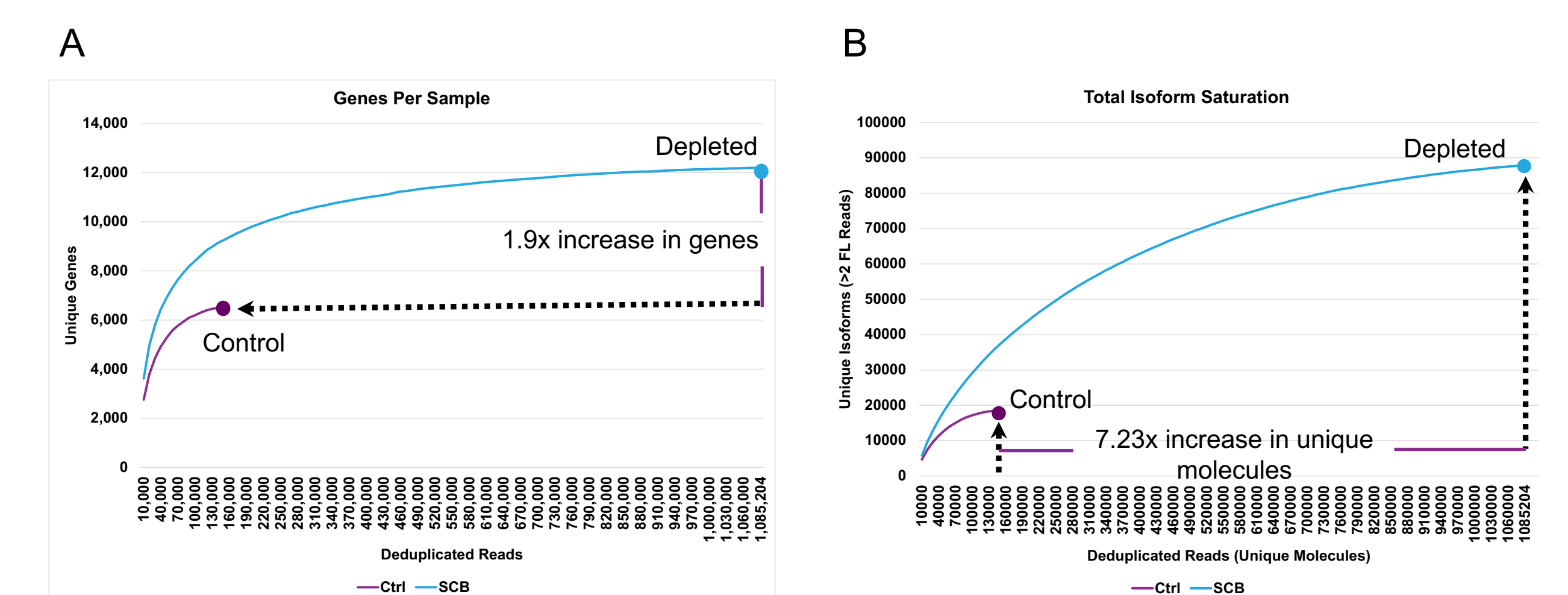
**Figure 3. IGV coverage view of reads aligning to two genes in the SCB + P800 guide module.** Binary alignment files from control and depleted samples were loaded into IGV in order to visualize the read coverage over two genes. A. The top IGV panel shows coverage over the TXNIP gene for the control (CTRL SCB + P800, purple) sample. The lower panel shows the depleted (DEPL SCB + P800, blue) sample. Read coverage over the TXNIP gene for the control sample is 0 - 4437 and 0 - 61 for the depleted. B. The top IGV panel shows coverage over the ACTB gene for the control (CTRL SCB + P800, purple) sample. The lower panel shows the depleted (DEPL SCB + P800, blue) sample. Coverage over the ACTB gene for the control sample is 0 - 53399 and 0 - 389 for the depleted. The average decrease in read coverage for both genes is approximately 100-fold.

### CRISPRclean enables isoform quantification and discovery

To assess the effect of CRISPRclean depletion on isoform discovery, libraries were depleted with and without CRISPRclean plus treatment. Again, multiple guide sets were tested independently and in combination. Figure 4 shows that the CRISPRclean method enables the discovery of more total and novel isoforms in depletions using the single cell boost (SCB) and single cell boost plus Panglao 800 (SCB + P800) guide modules. The increase of novel not in catalog isoform detection for these two modules was 345% and 311%, respectively. CRISPRclean depletion removes fragments prior to sequencing and enables more informative and lower expressed molecules to be sequenced.



**Figure 4. More isoforms are identified in CRISPRclean treated libraries than in untreated.** Total and novel isoform species were identified and counted in PacBio Iso-seq data and compared across control and depleted samples. A. Total isoforms species with at least two full length reads (FL Reads, y-axis) were calculated for control (CTRL) and depleted (DEPL) samples. All guide sets, coding ribosomal and mitochondrial (Ribo/Mito), single cell boost (SCB) and single cell boost plus Panglao 800 (SCB + P800) guides set enabled the discovery of more isoforms in depleted samples (Ribo/Mito = 231%, SCB = 375%, SCB + P800 = 354%). The number of novel in catalog (Novel IC = novel isoform with known junctions, blue) and not in catalog (Novel NIC = at least one novel junction, purple) isoforms were calculated for control (CTRL) and depleted (DEPL) samples across the three guide set modules (B - D). Depletion using all three guide modules showed an increase in novel and novel not in catalog isoform detected (Ribo/Mito = 197%, SCB = 345%, SCB + P800 = 311%).



**Figure 5. Depletion increases information gained and unique molecules in single cell IsoSeq samples.** Filtered reads from control (purple) and depleted (blue) PBMC samples were down sampled in 10,000 read bins (x-axis) and the number of unique genes (panel A) and isoforms (panel B) was calculated for each bin (y-axis). Data from the depleted sample contained 7.23-fold more unique molecules, 1.9-fold more unique genes and 4.5-fold more unique isoforms than the control. Further, this data demonstrates that more sequencing would not enable increased discovery in the control sample and that only depletion provides the ability to increase discovery metrics.

## 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. The CRISPRclean method helps to mitigate this effect by removing fragments produced by highly expressed genes, prior to sequencing, thus preserving sequencing space for informative molecules. The results presented here show:

1. Successful use of approximately 130,000 guides targeting 1699 genes.
2. Reads originating from targeted genes are depleted at > 90% in all guide modules tested.
3. CRISPRclean depletion results in higher numbers of isoforms with all three of the guide modules tested.
4. Novel not in catalog isoforms increased significantly with CRISPRclean depletion with all three guide modules.

To learn more visit [jumpcodegenomics.com](https://jumpcodegenomics.com)

1. Pettit 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, Felveson 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. Huang KK, Huang J, Wu JKL, Lee M, Tay ST, Kumar V, Ramnarayanan K, Padmanabhan N, Xu C, Tan ALK, Chan C, Kappai D, Gökçe 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.  
4. Veiga DFT, Nesta A, Zhao Y, Mays AD, Huynh R, Rossi R, Wu TC, Palucka K, Anczukow O, Beck CR, Bancheau J. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv. 2022 Jan 21;8(3).  
5. 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).