# Introduction to CRISPRclean®: the *in vitro* use of CRISPR-Cas to refine NGS applications

Azeem Siddique, Sridhar Ranganathan, Zenas George, Dhanya Ramachandran, Jon Bezney, Sonal Chaudhary, Gaia Suckow, Dante DeAscanis, Jeffrey Deason, Diana De La Toba, Yvain Desplat, Josh Diaz, Faizan Khalid and Jon Armstrong
*Jumpcode Genomics, San Diego CA*

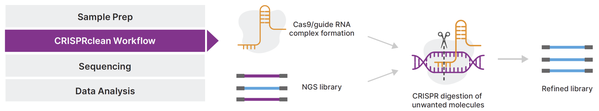
## Introduction

Next generation sequencing has replaced PCR and microarrays as the method of choice to gather information on the genomic or transcriptomic profile of an organism. Detection sensitivity and costs continue to be issues of concern when generating and evaluating sequencing data. To reduce costs, one might employ a targeted panel strategy, or sequence to low depth; to increase sensitivity, one might sequence to greater depth at the cost of sample number. Another method, far more powerful in its ability to enable discovery and minimize the trade-offs mentioned earlier is to remove known but unwanted sequences from NGS libraries prior to sequencing. Depletion, using probe-based hybridization methods, has been employed to remove ribosomal RNA (rRNA) from RNA-Seq libraries destined for sequencing, but the use of depletion-based approaches beyond rRNA has been limited.

We have employed the CRISPRclean workflow, which uses the CRISPR Cas9 double-stranded DNA endonuclease, to remove unwanted sequences from NGS libraries. The Cas9 enzyme targets DNA for cleavage in a site-specific manner when complexed with CRISPR RNA (crRNA). The latter directs Cas9 to a DNA sequence with complementarity to a 20-nucleotide target-specific sequence in the crRNA and Cas9 cleaves at that site. Thus, Cas9 can be used to target any known but unwanted sequence to remove it before it is sequenced. Such sequences could include ribosomal RNA sequences from individual species or mixed species samples, abundant but biologically uninformative sequences from single cell RNA-Seq libraries and human host DNA from metagenomic libraries. In each case, depletion results in increased sensitivity for sequences of interest and/or reduced cost of sequencing. Results of some applications are described here.

## CRISPRclean workflow

**Figure 1:** A simple, specific and effective method to remove unwanted and uninformative sequences from NGS libraries.




## Conclusions

- CRISPRclean is effective in removing uninformative sequences from NGS libraries.
- CRISPRclean increases sensitivity and lowers background noise to gain greater confidence in detecting differential gene expression.
- CRISPRclean Plus efficiently removed human and bacterial rRNA from fecal samples which resulted in 3-5 fold more bacterial species detected.
- Use of the CRISPRclean Single Cell RNA Boost kit with 10x Genomics Chromium libraries prepared from peripheral blood mononuclear cells (PBMCs) results in increased UMIs/cell and genes/cell and a higher biological signal to noise ratio. (See Poster 546 from Jumpcode Genomics for more details).

## Library prep with integrated CRISPR-based rRNA depletion for total RNA-Seq workflows

**CRISPRclean Stranded Total RNA Prep with rRNA Depletion Kit (Human, Mouse, Rat) produces high quality rRNA-depleted libraries from several species.**

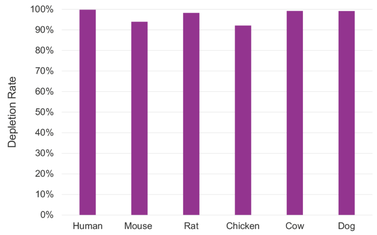



**Figure 2a:** CRISPRclean shows effective depletion of rRNA from multiple species.

Libraries prepared from human, mouse, rat, dog, cow and chicken total RNA (Zyagen) were depleted with CRISPRclean. Depletion rates were calculated by aligning reads to the rRNA genes obtained from a high-quality reference genome of each species.

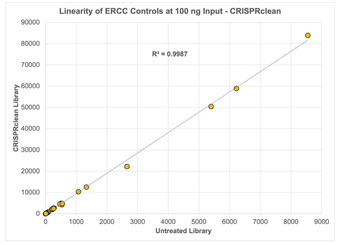



**Figure 2b:** CRISPRclean depletion has little effect on ERCC control RNA.

CRISPRclean libraries were prepared in triplicate from 100 ng of Universal Human RNA (UHR) spiked with ERCC synthetic control RNA. A comparison of ERCC read counts between untreated and CRISPRclean depleted libraries is shown here.

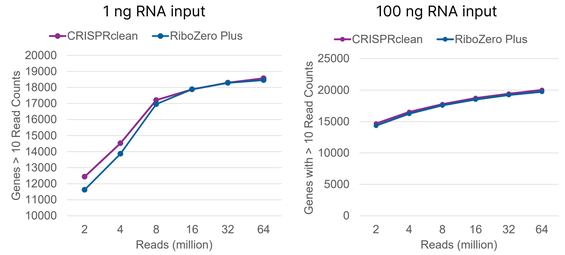



**Figure 2c:** Library complexity of CRISPRclean is comparable to Illumina Stranded Total RNA libraries with Ribo-Zero Plus.

CRISPRclean libraries prepared from 1 ng and 100 ng UHR inputs were compared to Illumina Total Stranded RNA-Seq libraries with RiboZero Plus depletion. Library complexity was calculated by down-sampling the data and counting the number of human protein coding genes with greater than 10 counts per million reads at each down-sampled read number.

## Increased bacterial species detection in microbiome stool samples

**CRISPRclean Plus Stranded Total RNA Prep with rRNA Depletion (Human, Mouse, Rat, Pan-Bacteria) is a library preparation method suited to the retrieval of transcriptomic information from complex multi-species samples. The product enables the simultaneous evaluation of RNA virus genomic data, microbiome composition, and host gene expression from complex biological samples, such as fecal, saliva, and nasal samples, that often contain a mixture of human and bacterial cells.**

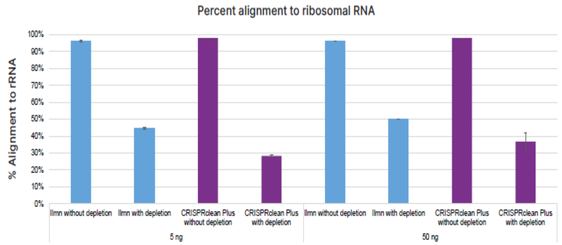



**Figure 3a:** CRISPRclean Plus is efficient at removing rRNA from mixed species samples.

CRISPRclean Plus libraries and Illumina Stranded Total RNA-Seq libraries with RiboZero Plus libraries were prepared in triplicate from 5 ng and 50 ng of human fecal RNA. Libraries were sequenced on a NovaSeq 6000 instrument. Depletion rates were calculated by aligning reads to rRNA sequences from the Silva and rfam rRNA databases.

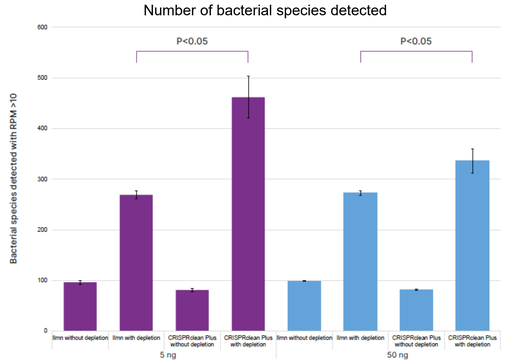



**Figure 3b:** Compared to Illumina RiboZero, CRISPRclean Plus depletion facilitates detection of more bacterial species.

Bacterial species were identified using Kraken2. A minimum threshold of 10 reads per million was used when calculating species numbers.

## Boost in usable single cell data by reducing wasted sequencing

**CRISPRclean Single Cell RNA Boost increases the biologically relevant output from 10x Genomics Chromium 3' gene expression libraries. It removes reads derived from rRNA, 100 ribosomal and mitochondrial protein coding genes and 155 genes with gene expression patterns that vary little between cell types, as well as reads derived from genomic intervals. By removing uninformative reads, CRISPRclean enables the redistribution of reads to more biologically relevant transcripts.**

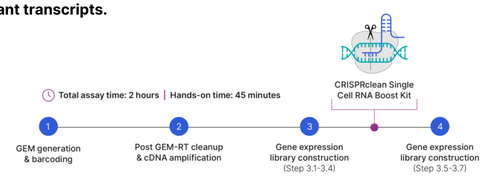



**Figure 4a:** Simple 3-step protocol integrated into 10x Genomics Chromium workflow.

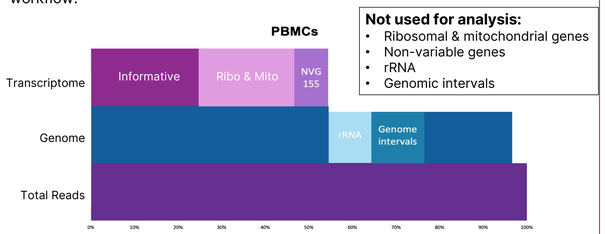



**Figure 4b:** CRISPRclean Single Cell Boost is designed to deplete sequences not used in secondary analysis, including ribosomal, mitochondrial and non-variable gene sequences, and intergenic reads.

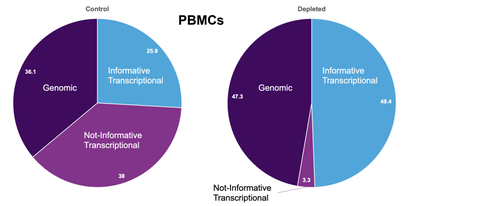



**Figure 4c:** 100% increase in reads mapping to the transcriptome using CRISPRclean Single Cell RNA Boost with 10x Chromium 3' gene expression libraries.

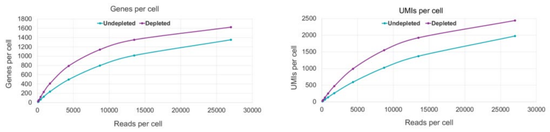



**Figure 4d:** Improvement in detection of genes/cell and UMIs/cell with depletion at all sequencing depths.



To learn more visit jumpcodegenomics.com | Download the poster