Resolving clinically relevant short read deficient homologous sequences using a novel CRISPR-CAS mediated targeted long read sequencing method.

Ma Z[1], Kothandaraman A[1], Balciuniene J[1], Guo F[1], Brown K[2], Siddique A[2], Ranganathan S[2], Armstrong JR[2], Hegde M[1]

[1]Revvity, Waltham, MA
[2]Jumpcode Genomics, San Diego, CA

Introduction

Short read genome and exome sequencing (GS/ES) along with advanced bioinformatic tools have been widely adopted in clinical settings due to the significant advantages over Sanger sequencing. Despite the advancements there are still numerous genomic regions of clinical utility that cannot be resolved due to high sequence homology with other regions of the human genome. Longer read lengths and paired end reads have proven helpful but remain incomplete. Synthetic long reads, optical mapping, Hi- C, along with brute force long range PCR and sanger sequencing have been able to resolve some of these evolutionary artifacts but are impractical for most clinical laboratories. A cost-effective solution for targeted long read sequencing is needed. Here we present work towards a novel long read targeted sequencing method that combines the in-vitro use of the CRISPR-Cas9 system with highly accurate PacBio sequencing using high molecular weight gDNA.

Figure 1a

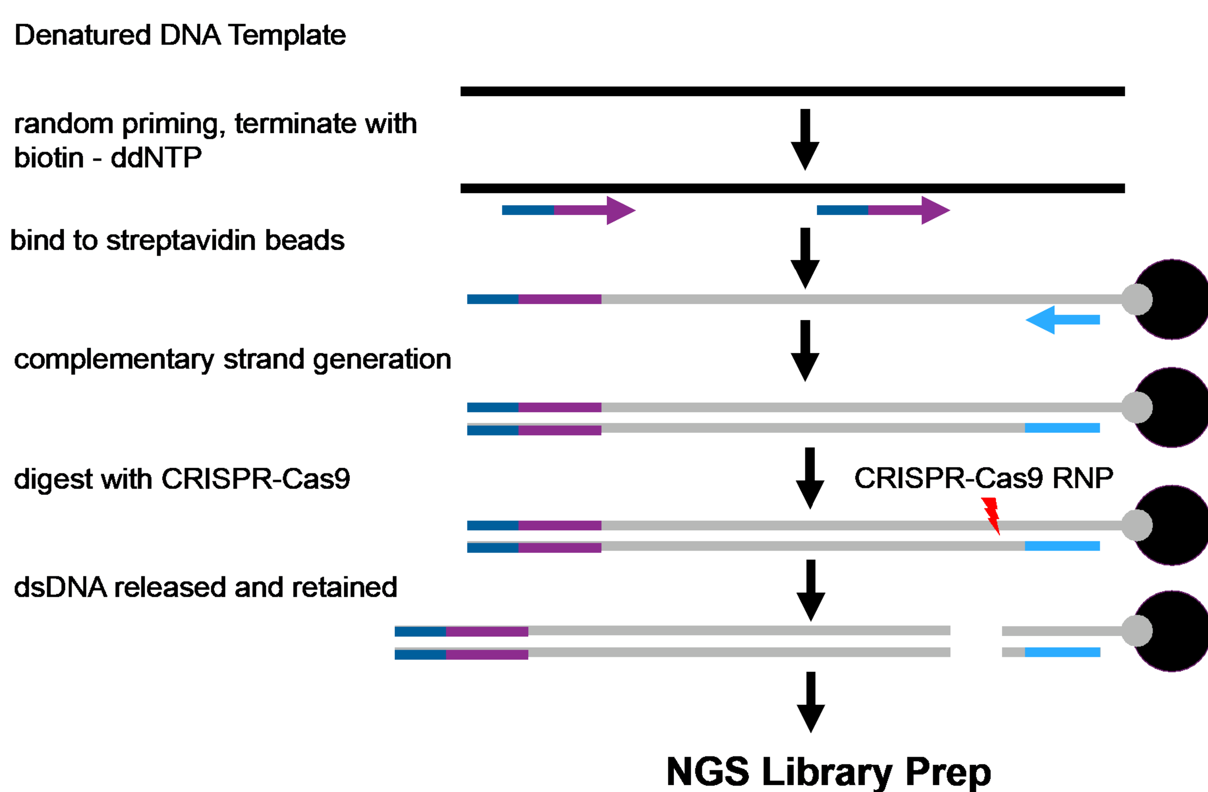
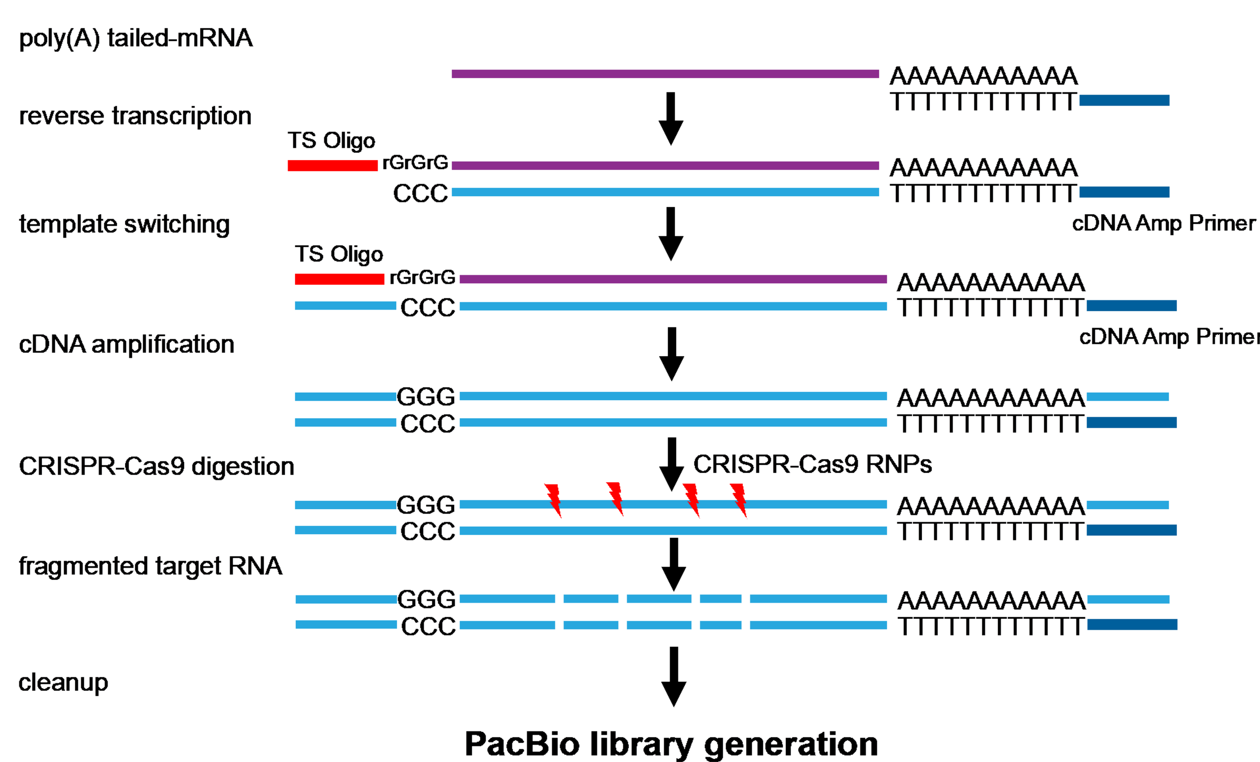


Figure 1b



Methods

On bead enrichment using targeted CRISPR-Cas9 cleavage. High molecular weight gDNA was obtained from an *Escherichia coli* (E. coli) isolate (ATCC). Extension products were generated using barcoded random primers. Extension products were terminated with biotinylated dideoxynucleotides (ddNTPs). The biotinylated extension products were captured on streptavidin coated magnetic beads and double stranded using a second random primer. 24 CRISPR-CAS9 guide RNAs were designed to target unique sequences flanking the E. coli MinCDE locus (5 kb region). Incubation of DNA products with Cas9 protein bound to guide RNA produced double stranded cuts at the guide RNA target sites. The high specificity of the CRISPR-CAS system ensures that only genomic fragments of interest are cleaved from the beads while non-target sequences remain bound. The target molecules released by CRISPR cleavage were eluted from the beads and underwent standard Illumina library construction. Sequencing was performed on a Illumina MiSeq with 2 x 150 bp paired-end reads. A workflow diagram of the method is shown in Figure 1a.

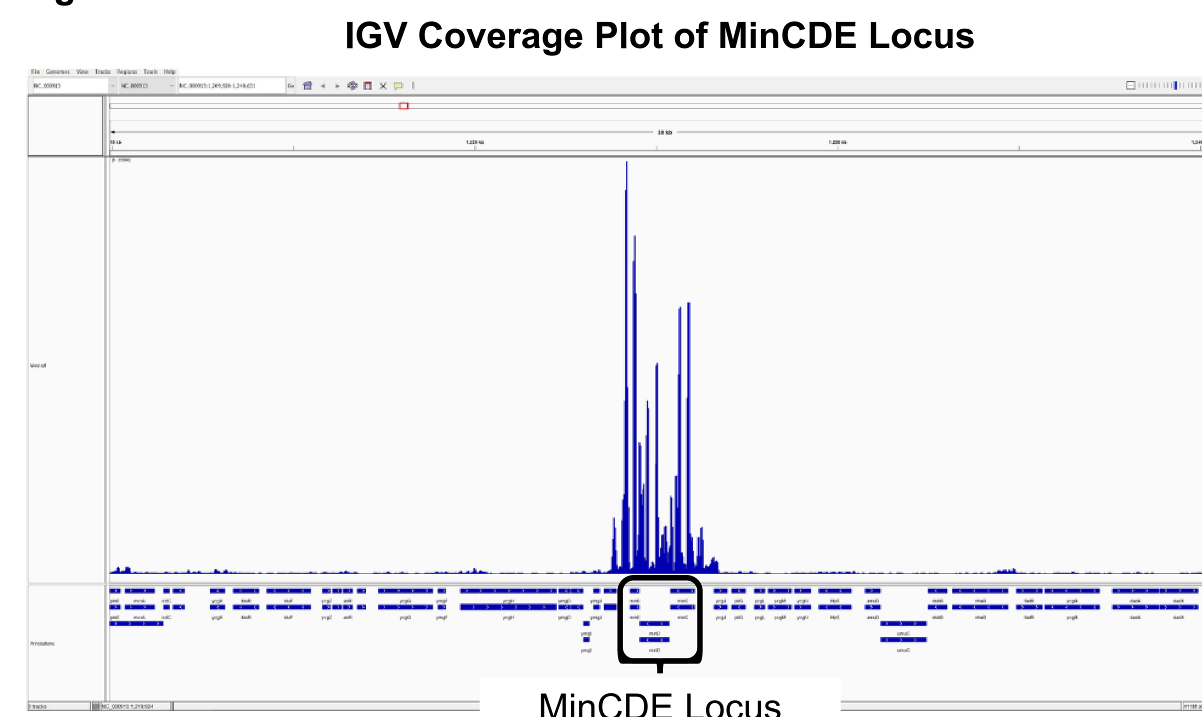
Depletion of long cDNA fragments using targeted CRISPR-Cas9 digestion. Total RNA was extracted from human PBMC samples and cDNA was generated using the Chromium Single Cell 3' library platform (v3.1). CRISPR-CAS9 guide RNAs were designed to target the highest expressed 90 human coding ribosomal and 10 mitochondrial RNAs. Incubation of cDNA products with Cas9 protein bound to guide RNA produced double stranded cuts at the guide RNA target sites. The cleaved fragments were removed using a standard magnetic bead cleanup and the remaining fragments were amplified by PCR. PacBio sequencing libraries were generated from the amplification products and sequenced on the PacBio Sequel system. A workflow diagram of the method is shown in Figure 1b.

Results

On bead targeted CRISPR-Cas9 cleavage highly enriches for the region of interest.
Figure 2 shows an IGV coverage plot of the E. coli MinCDE locus. Read coverage over the MinCDE locus reaches a level of 20,000-fold with low background coverage of the surrounding areas.

Figure 2

IGV Coverage Plot of MinCDE Locus



CRISPR-Cas9 is effective at depleting long cDNA fragments.
Figure 3a shows a histogram of PacBio sequencing read counts (y-axis) binned by reads length (x-axis) for untreated (blue) and treated (orange) sample Sc-A. A shift is observed in the library size of the treated sample due to CRISPR-Cas9 targeted removal of coding ribosomal and mitochondrial RNA transcripts from the library. The effect of this removal is seen in the Figure 3b where the combined percent of reads (y-axis) aligning to these gees is 15.96%-17% in the control and 0.14%-0.12% in the depleted samples Sc-A and Sc-B, respectively. This represents a depletion rate of 99.2%.

Figure 3a

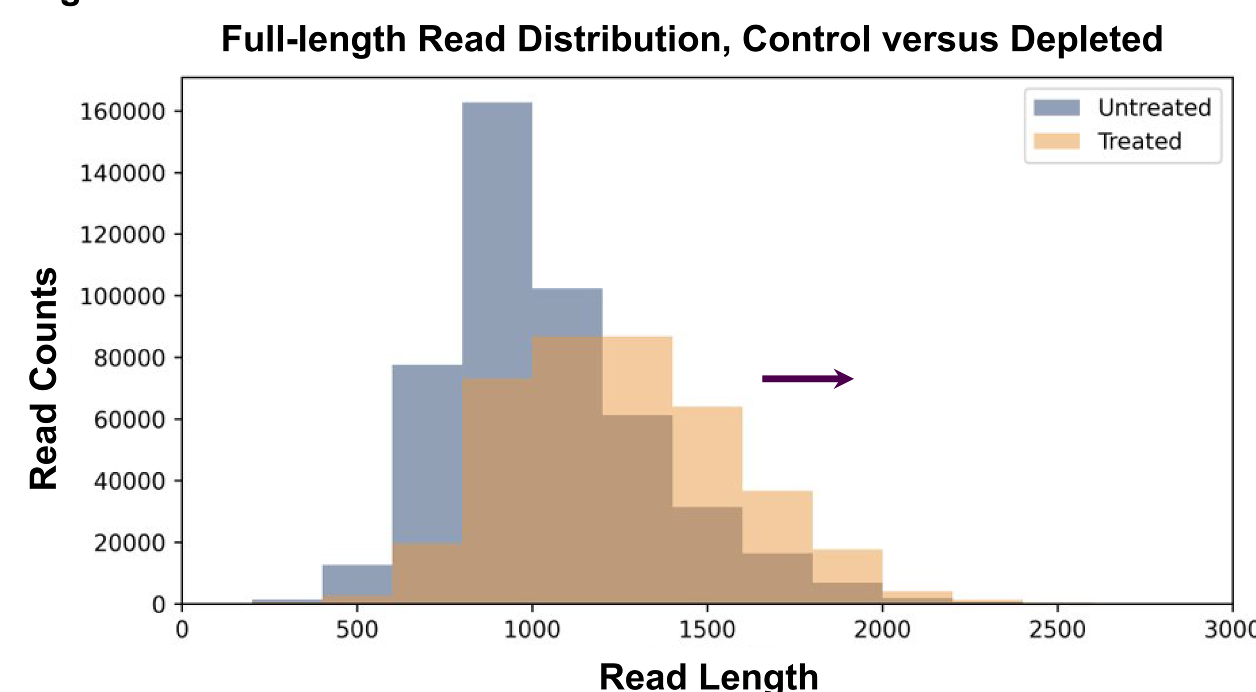
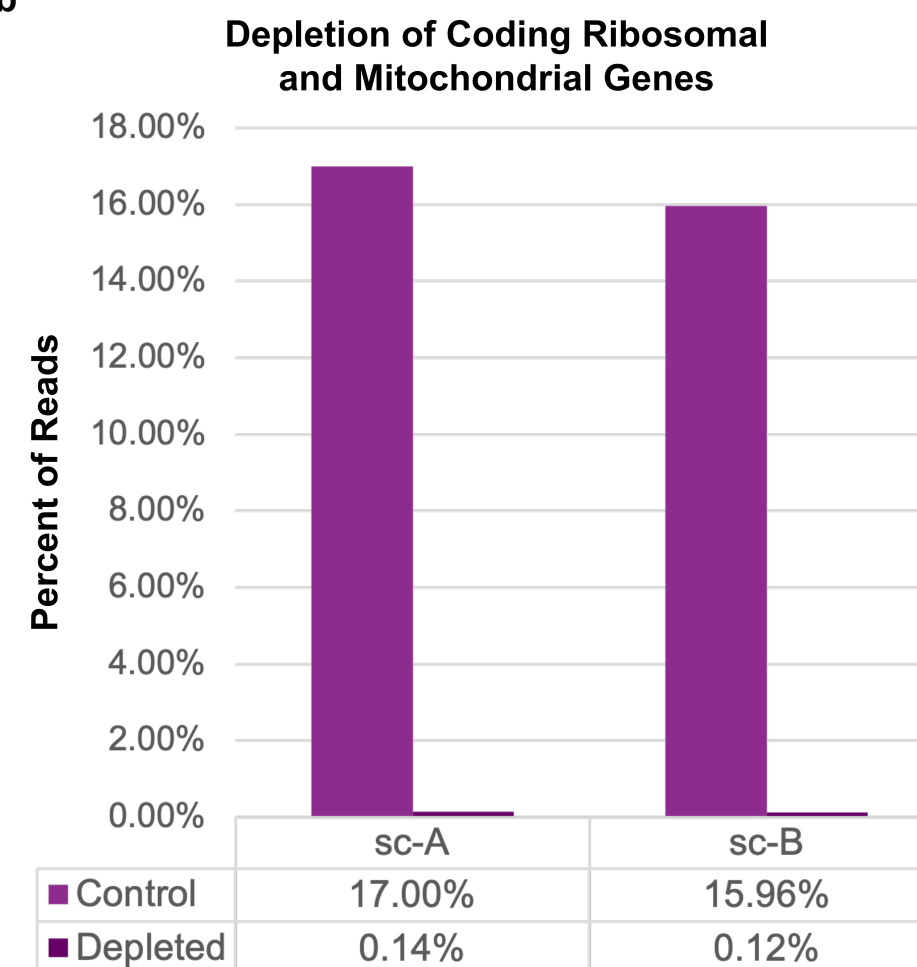


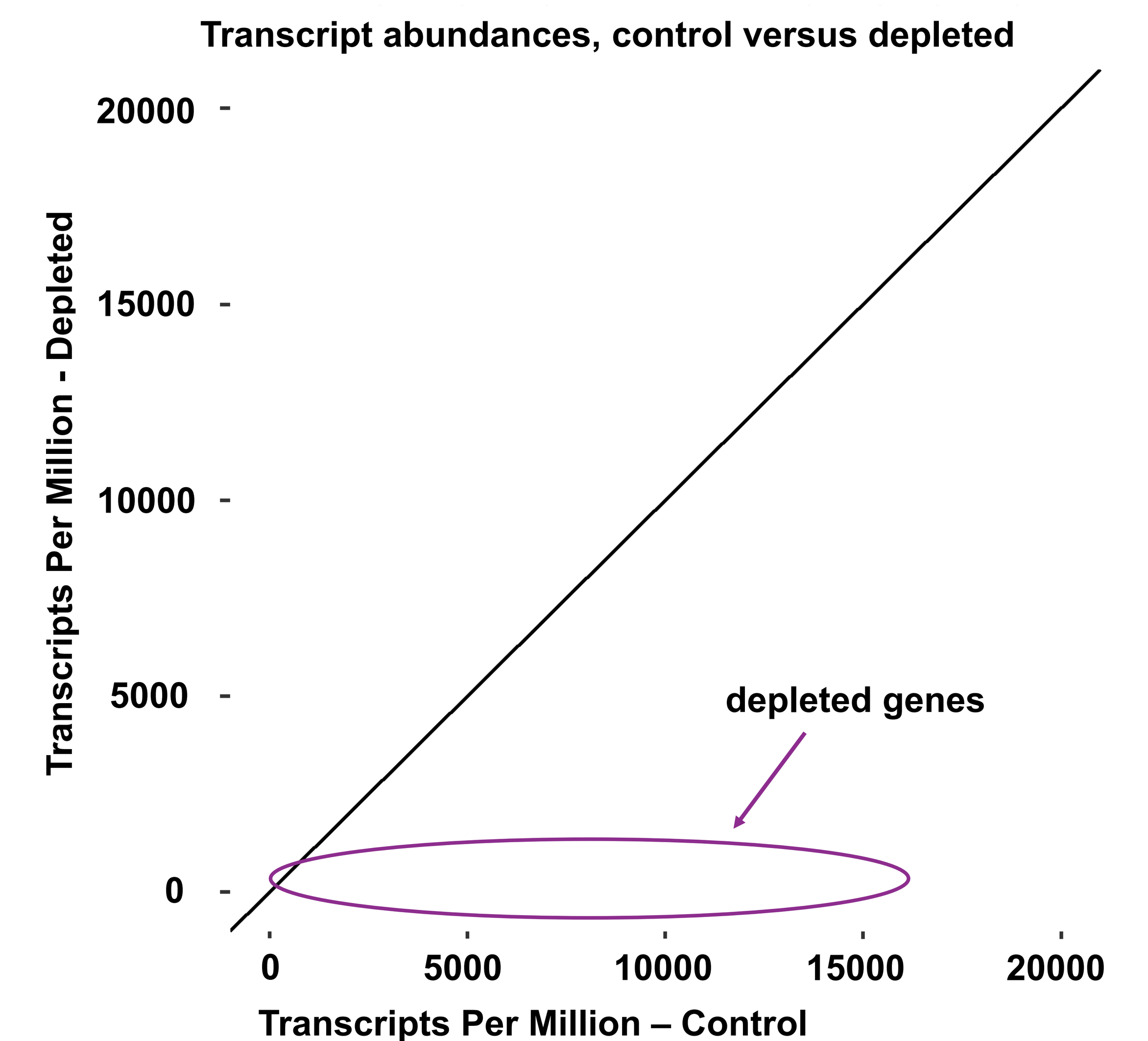
Figure 3b



Average 99.2% depletion rate

CRISPR-Cas9 depletion enables effective depletion of large transcript fragments with low bias.
PacBio sequencing reads were generated for depleted and control samples and the transcripts per million were calculated per gene. Figure 4 shows the comparison of gene expression levels between control (x-axis) and depleted (y-axis) samples. The graph clearly shows effective depletion of targeted coding ribosomal and mitochondrial transcripts. In addition, depletion exhibits low bias between control and depleted samples.

Figure 4



Conclusions

- CRISPR-Cas9 cleavage and enrichment can be applied to fragments bound to beads.
- The application of CRISPR-Cas9 cleavage to fragments bound to beads provides the ability to highly enrich for regions of interest.
- The programmability of the CRISPR-Cas9 system, enables the design of guides to enrich for clinically relevant genes with complex homologous areas in the genome.
- CRISPR-Cas9 digestion is highly effective on long fragments.
- The combination of these methods would provide a novel approach for enrichment of clinically relevant genes with complex homology while generating long read sequences that provide full length information across the gene.

Future Work

Future work will focus on combining the two methods presented into a single workflow that can be sequenced on short or long read sequencing systems.

revvity JUMPCODE GENOMICS

© Copyright 2023, Jumpcode Genomics, Inc.; all rights reserved