CRISPR-MEDIATED DEPLETION OF RIBOSOMAL RNA-DERIVED DNA FROM RNA-SEQ NGS LIBRARIES

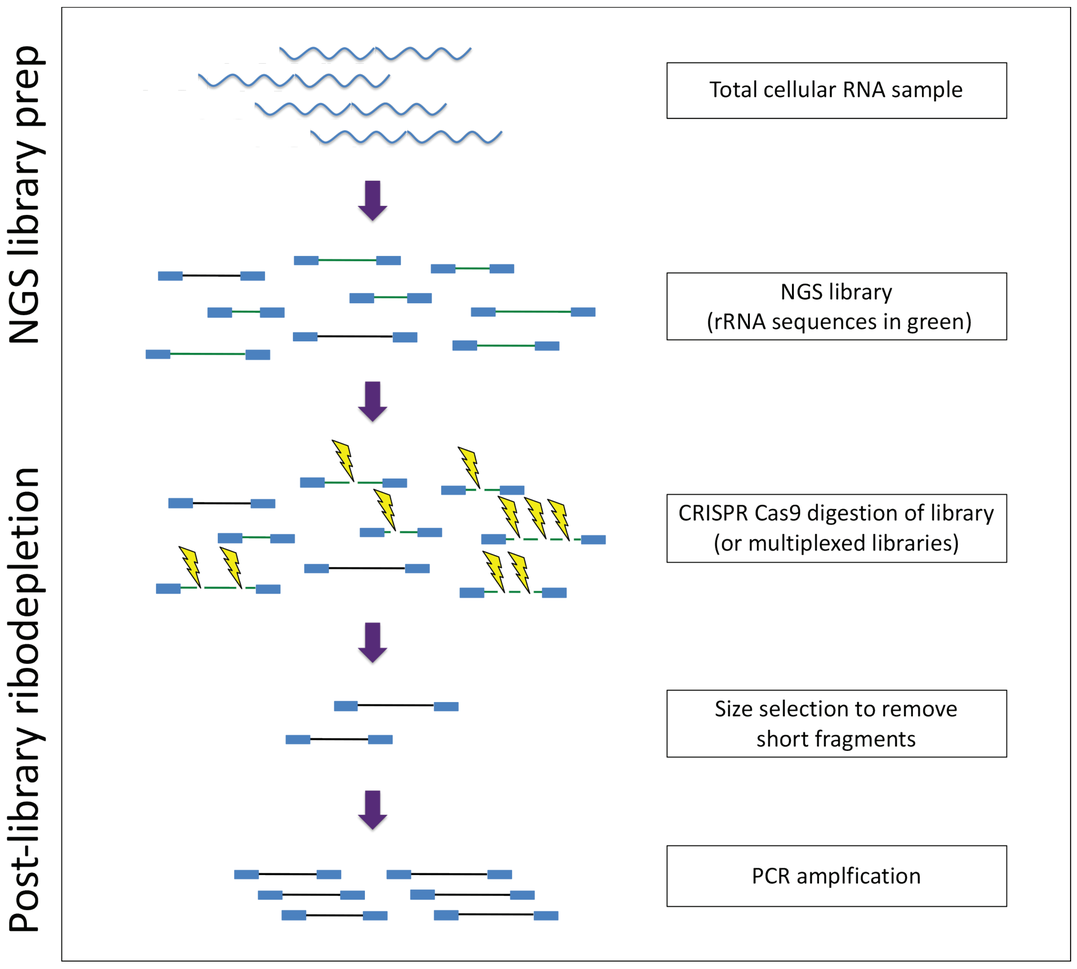
Azeem Siddique,[1,3] Gaia Suckow,[1,3] Hallie Sussman,[1,3] Dhanya Ramachandran,[1,3] Nils Homer,[2] Jorge Bahena,[1,3] Phillip Ordoukhanian,[1,3] Steve Head,[1,3] Keith Brown[3] | [1]The Scripps Research Institute, La Jolla, CA; [2]Fulcrum Genomics, Somerville, MA; [3]JumpCode Genomics, Carlsbad, CA

Abstract

Although messenger RNA (mRNA) is the focus of much RNA research, it constitutes a relatively small fraction of total RNA in a cell. Most cellular RNA is ribosomal RNA (rRNA) and removal of this RNA is is desirable in many RNA-Seq studies in order to maximize capacity of the sequencing instrument and reduce costs. There are a number of methods to achieve this: bead-bound poly-T oligonucleotides can be used to enrich for poly-A tailed mRNA, or rRNA can be removed by hybridization to short complementary sequences and capture by beads or subsequent RNase H treatment. However, rRNA removal methods tend to be expensive and often require large amounts of RNA as input.

We have employed the CRISPR Cas9 double-stranded DNA endonuclease to develop a different method of rRNA removal for RNA-Seq studies. The CRISPR Cas9 enzyme targets DNA for cleavage in a site-specific manner when it is 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. In this CRISPR-mediated rRNA depletion method, a library of crRNAs are designed based on targets in the abundant classes of rRNA genes. Completed RNA-Seq libraries produced from total cellular RNA are treated with Cas9 complexed to the library of crRNA guides. NGS library molecules containing rRNA sequences are cleaved and, thus, are no longer substrates for PCR or sequencing. After Cas9 treatment, the library is size selected, PCR amplified and sequenced. Here we show data from the CRISPR treatment of rRNA sequences from RNA-Seq libraries prepared from three different bacteria - *Escherichia coli*, *Methylococcus capsulatus* and *Staphylococcus aureus* - and *Homo sapiens*.

Schematic of the protocol for CRISPR-mediated rRNA depletion



Conclusions

- CRISPR technology can be used to remove unwanted DNA from next generation sequencing libraries.
- Greater than 99% of reads derived from ribosomal RNA were removed after CRISPR treatment using CRISPR guide RNA sets designed from the rRNA sequences of four different organisms.
- The technology can be customized to any organism for which a reasonably high quality reference genome exists.
- Because CRISPR-mediated targeted DNA depletion occurs after the libraries has been prepared, the technology has the potential to be employed on multiplexed libraries.

Ribosomal RNA depletion metrics for three bacteria

Samples	*E. coli* (untreated)	*E. coli* (Cas9 treated)	*S. aureus* (untreated)	*S. aureus* (Cas9 treated)	*M. capsulatus* (untreated)	*M. capsulatus* (Cas9 treated)
Genome alignment rate	98.4%	98.2%	99.1%	96.9%	98.6%	91.3%
Duplication rate	40.0%	11.7%	37.3%	10.0%	61.6%	7.6%
Alignment rate (16S rRNA)	37.9%	0.2%	27.5%	0.1%	36.1%	0.1%
Alignment rate (23S rRNA)	60.9%	0.5%	65.3%	0.2%	59.9%	0.2%
Alignment rate (cds)*	1.0%	93.1%	6.1%	87.6%	3.6%	88.0%
Mean coverage of cds*	0.2X	15.3X	1.7X	23.8X	1.6X	36.5X
rRNA depletion rate	-	99.3%	-	99.8%	-	99.6%

* cds = coding sequences (coverage calculated at 1 million reads for *E. coli* and *S. aureus* and 2 million reads for *M. capsulatus*)

The NEBNext Ultra II Stranded RNA-Seq Library Prep Kit was used to prepare an NGS library from 100 ng of RNA from *Escherichia coli*, *Staphylococcus aureus* and *Methylococcus capsulatus*. One ng of each completed library was treated with CRISPR Cas9 pre-complexed with a library of CRISPR RNA guides. The workflow described in this poster was used to deplete the library of ribosomal RNA sequences. The non-depleted and depleted libraries of each organism were sequenced on an Illumina MiSeq 2 x 75 bp flow cell. The online software, CRISPOR (www.crispor.tefor.net), was used to design guides to the rRNA sequences (5S, 16S & 23S) of each organism. The RNA samples were obtained from Thermo Fisher (*Escherichia coli*) or kindly provided by Intrexon (*Methylococcus capsulatus*) and the Steinke laboratory at the University of Nebraska (*Staphylococcus aureus*).

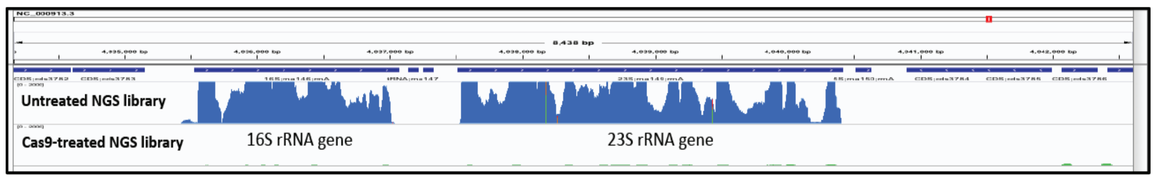
Ribosomal RNA depletion metrics for *Homo sapiens*

Samples	Genome alignment rate	Duplication rate	Alignment rate (18S rRNA)	Alignment rate (28S rRNA)	Alignment rate (12S rRNA)	Alignment rate (16S rRNA)	Alignment rate (exon)	Alignment rate (intron)	Alignment rate (intergenic)	rRNA depletion rate
Homo sapiens (untreated)	99.3%	52.5%	11.4%	22.7%	1.2%	1.4%	3.7%	3.2%	15.9%	-
Homo sapiens (Cas9 treated)	99.3%	4.0%	0.0%	0.1%	0.0%	0.0%	48.5%	29.6%	13.6%	99.0%

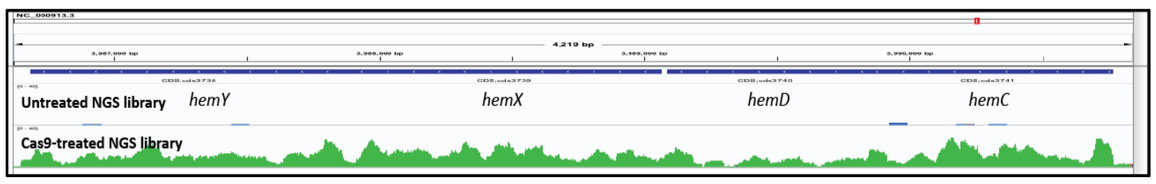
The NEBNext Ultra II Stranded RNA-Seq Library Prep Kit was used to prepare an NGS library from 100 ng of total human brain RNA (Thermo Fisher). One ng of the completed library was treated with CRISPR Cas9 pre-complexed with a library of CRISPR RNA guides, according to the protocol outlined in this poster. The CRISPR RNA guides are designed to target the human nuclear and mitochondrial ribosomal RNA genes. CRISPOR software was used to design guides to the nuclear rRNA (5S, 5.8S, 18S & 28S) sequences while custom software built by Fulcrum Genomics was used for the mitochondrial rRNA (12S & 16S) guide design.

Read coverage with and without ribodepletion at two loci in *Escherichia coli*

rRNA locus



hemCDXY locus



Illumina sequence data from rRNA-depleted and non-depleted *Eschericiha coli* total RNA libraries were viewed in the Integrated Genome Browser (IGV 2.3). Coverage profiles of the two samples are shown here for two loci: the *rRNA* operon (top) and *hemCDXY* locus (bottom). The data shows depletion of *rRNA* sequence-containing reads and enrichment of *hemCDXY* sequence-containing reads in the CRISPR Cas9-treated library. All sequence data were downsampled to one million reads prior to visualization.