JUMPCODE GENOMICS

TECH NOTE

CRISPRclean® depletion technology: A new tool to better understand rare diseases

Enhanced RNA-Seq delivers double the genes detected with confidence

Introduction

Current first-line genetic tests for rare diseases include chromosomal microarray (CMA) and clinical exome sequencing. However, these methods are insufficient for a large proportion of cases. Whole genome sequencing (WGS) is commonly used to detect rare structural variants and small copy number variations (CNVs). Combining RNA sequencing (RNA-Seq) with WGS promises to enhance our understanding of rare diseases' mechanisms.[1-2] This approach allows researchers to observe the aberrant transcription of RNA from DNA by allowing the calculation of the splicing ratio for novel unannotated junctions and reporting the proportion of reads that support aberrant splicing. Significantly, RNA-Seq increases the detection of low-expressing transcripts and isoforms for rare diseases.[3-7]

Typically, most genes in Online Mendelian Inheritance in Man (OMIM) are expressed in accessible tissues from any given patient, though at low levels and cannot be used for splicing analysis. Many RNA-Seq reads originate from abundantly expressed genes and are irrelevant to a given rare disease etiology. However, it is possible to remove this abundant noise before sequencing by targeting genes with transcripts per million (TPM) > 30. Removing these high-expression genes results in sequencing reads being redistributed to low and medium-expressed genes, providing a more confident sampling of this minimally observed part of the transcriptome. The increased coverage significantly boosts the detection sensitivity and confidence of aberrant gene expression signals. Here we show that using CRISPRclean technology, abundant and uninformative RNAs can be easily removed from NGS libraries. This approach leverages the in vitro depletion of library fragments with CRISPR/Cas9 complexes programmed with guide RNAs to desired targets to be removed. The application of CRISPRclean technology to fibroblast samples doubles the number of observed genes at TPM > 30 because of increased coverage from reassigned reads, thereby increasing the reliability of analysis and expanding discovery potential.

Methods

Three fibroblast total RNA samples were split into two aliquots and depleted in triplicate using ~450,000 guides designed against 4,326 high-expression genes (> 30 TPM). Guides were designed to target 70% of the total raw reads for depletion. Control (non-depleted) samples were retained in triplicate from the second aliquot. Sequencing libraries were generated from control, and post-depletion replicates using the NEBNext® Ultra II Stranded Total RNA with Poly A module kit. Size selection of the final library was performed at 400-500 bases, and libraries were sequenced on the NextSeq® 2000 as 2×150 base reads. All samples were normalized to 100 MM reads for analysis.

Results

Depletion significantly increases the coverage of low and medium-expression genes

The sequencing results of the fibroblast samples show a significant boost in coverage for low and medium-expression genes for the depleted samples when compared to the control. In Figure 1, the purple population of genes is the low to medium-expression genes in the CRISPRclean depleted condition, which is now covered at TPM > 30. These genes are not covered at that level in the control sample. Genes with TPM >

30 in both the control and depleted samples are shown in blue. 4,477 genes were raised above a threshold of 30 TPM due to depletion in the CRISPRclean treated sample. This represents 4,477 genes that now have sufficient coverage for rare disease researchers to proceed with further analysis reliably.

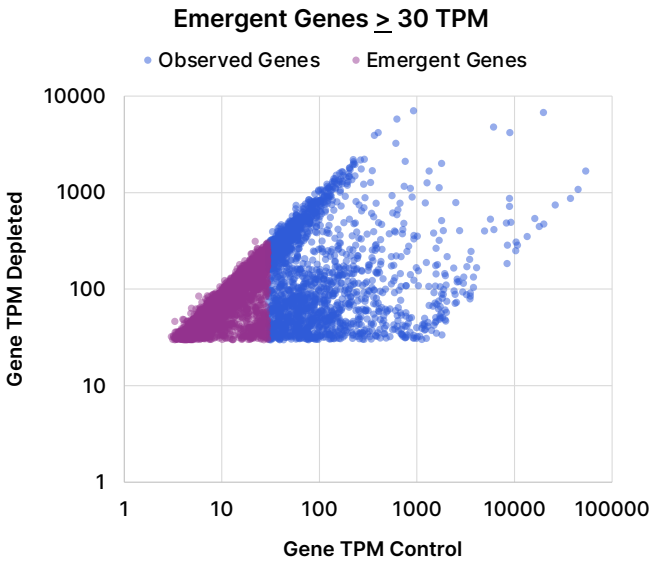


Figure 1: Depletion increases the number of genes above 30 transcripts per kilobase million (TPM). The number of genes above 30 TPM was calculated for a control (x-axis) and depleted (y-axis) fibroblast sample. All points represent the average of all sample replicates (control – n=9, depleted – n=9

Depletion doubles the number of genes with TPM >30 in fibroblast samples.

Further, in Figure 2, all genes with a TPM of greater than 30 are shown in a Venn Diagram for both the control and depleted sample. The overlap between the two populations is only 1,951 because of the removal of the targeted, well-characterized genes in the control population and the subsequent reallocation of reads onto lower expression genes in the depleted sample. Additionally, the 4,477 genes brought over the TPM of greater than 30 in the depleted sample include 3380 genes OMIM genes. Of the 1429 genes that were only in control samples, 1336 were targeted for depletion and predicted to decrease in the depleted samples. When comparing the full population of genes with a TPM of greater than 30 in the control to the depleted sample, there are 3380 genes in the control population and 6425 genes in the depleted sample. Thus, using CRISPRclean depletion to boost low and medium-expression genes doubles the discovery space for the fibroblast samples.

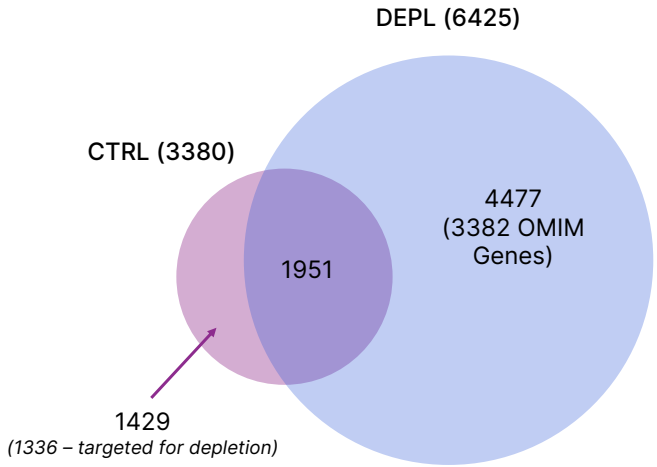
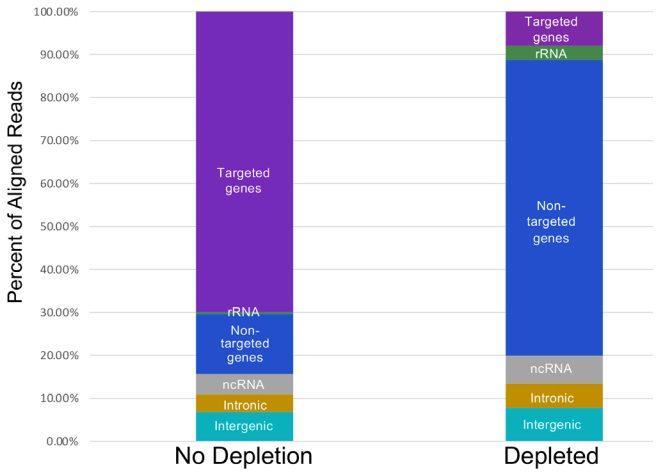


Figure 2: Depletion doubles the number of genes now detectable above TPM of >30 compared to control in fibroblast samples. The graph represents the average of all sample replicates (control – n=9, depleted – n=9)

Depletion shows ~5-fold increase in reads aligning to non-targeted genes

Finally, in Figure 3, the five-fold change in non-targeted genes can be seen in the depleted sample compared to the control sample (dark blue). In the control sample, the targeted genes occupy nearly 70% of the sequencing space; in the depleted sample, this shrinks to 13% (purple). Removing these reads allows the non-targeted genes to occupy more sequencing space, thus boosting them over the 30 TPM threshold and allowing novel splice sites in low abundance to obtain greater sequencing coverage. Depletion increased the percent of reads aligning to non-targeted genes in depleted samples by ~5-fold, while the percent of reads aligning to targeted genes decreased to 8.01%.



	No Depletion	Depleted
intergenic	6.83%	7.73%
intronic	4.07%	5.68%
ncRNA	4.73%	6.51%
rRNA	0.58%	3.25%
targeted genes	69.83%	8.01%
non-targeted genes	13.94%	68.81%

Figure 3: Depletion shows ~5-fold increase in reads aligning to non-targeted genes.

Conclusion

Molecular depletion significantly increases coverage of low abundance species in RNASeq. By depleting nearly 60% of uninformative and abundantly expressed genes, we confidently detected an additional 3,382 OMIM genes in the depleted samples, an increase of 230%. Importantly, these genes were not detectable at a 30 TPM threshold in control samples. CRISPRclean depletion is an effective tool to increase discovery potential for transcriptomic researchers looking to confidently investigate parts of the transcriptome that have been obscured by abundant and uninformative gene expression.

To learn more, visit jumpcodegenomics.com

Ordering information

Catalog	Product name	Samples
KIT1020	CRISPRclean® High Expressing RNA Depletion Kit	24

References

1. Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O'Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., Gonorazky, H., ... MacArthur, D. G. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science translational medicine*, 9(386), eaal5209. https://doi.org/10.1126/scitranslmed.aal5209
2. Kremer, L., Bader, D., Mertes, C. *et al.* Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8**, 15824 (2017). https://doi.org/10.1038/ncomms15824
3. Splinter, K., Adams, D. R., Bacino, C. A., Bellen, H. J., Bernstein, J. A., Cheatle-Jarvela, A. M., Eng, C. M., Esteves, C., Gahl, W. A., Hamid, R., Jacob, H. J., Kikani, B., Koeller, D. M., Kohane, I. S., Lee, B. H., Loscalzo, J., Luo, X., McCray, A. T., Metz, T. O., Mulvihill, J. J., ... Undiagnosed Diseases Network (2018). Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *The New England journal of medicine*, 379(22), 2131–2139. https://doi.org/10.1056/NEJMoa1714458
4. Byron, S., Van Keuren-Jensen, K., Engelthaler, D. et al. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet* **17**, 257–271 (2016). https://doi.org/10.1038/nrg.2016.10
5. Yépez, V.A., Gusic, M., Kopajtich, R. *et al.* Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med* **14**, 38 (2022). https://doi.org/10.1186/s13073-022-01019-9
6. Marco-Puche, G., Lois, S., Benítez, J., & Trivino, J. C. (2019). RNA-Seq Perspectives to Improve Clinical Diagnosis. Frontiers in genetics, 10, 1152. https://doi.org/10.3389/fgene.2019.01152
7. Peymani, F., Farzeen, A., & Prokisch, H. (2022). RNA sequencing role and application in clinical diagnostic. Pediatric investigation, 6(1), 29–35. https://doi.org/10.1002/ped4.12314