

# A Negative Result on Exome Sequencing: What a Genetic Counselor Should Know



Gemma Chandratillake, Sarah Garcia, Jeanie Tirch, Deanna M. Church, Mark Pratt, Michael J. Clark, Jason Harris, Stephen Chervitz, John West, Richard Chen  
Personalis, Inc. | 1350 Willow Road, Suite 202 | Menlo Park, CA 94025

## Introduction

Exome sequencing is increasingly utilized in the diagnosis of disorders of unknown genetic etiology. Despite increasing evidence for the utility of this approach, the majority of patients presenting with a presumed Mendelian genetic trait receive a negative sequencing report. Both technical and biological factors contribute to the current difficulty in deriving diagnoses from genomic sequence information, and genetic counselors ordering exome sequencing tests should be aware of the following issues.

## Whole Exome Sequencing is Not as Comprehensive as it Sounds

### Exome Sequencing Misses Exons

Many genes with clinical relevance are not well-covered by standard exome sequencing (FIGURE 1A). Exome enrichment kits do not provide coverage over all exonic content; several genes are completely absent from exome sequences, and many more genes are only partially covered. Reasons for poor coverage include lack of targeting by exome platforms, and high guanine/cytosine content, which is particularly prevalent in first exons.

FIGURE 1A: Coverage plot of *RPGR* using a standard exome sequencing platform

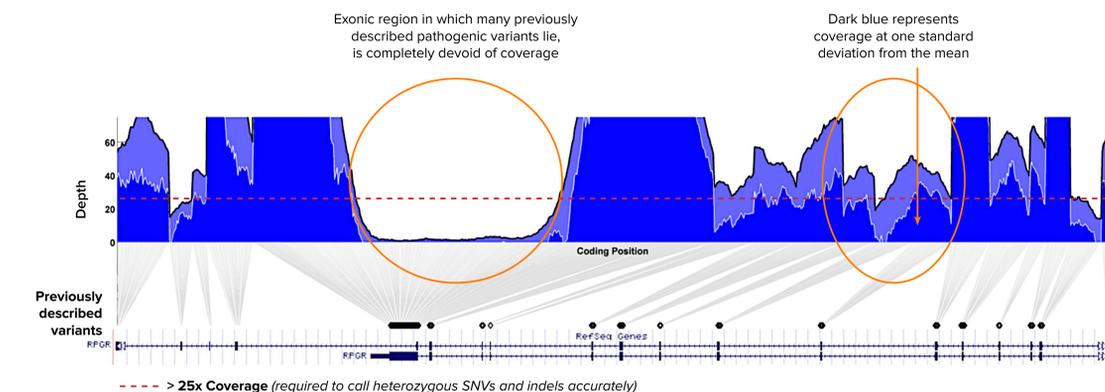
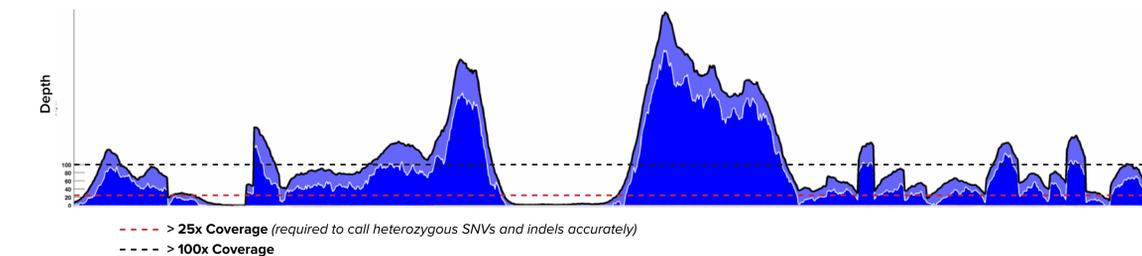


FIGURE 1B: High depth coverage plot of *RPGR* using a standard exome sequencing platform



### Average Coverage Statistics Do Not Necessarily Reflect Good Gene Coverage

"Average depth" or "mean coverage" statistics are often compared when discussing coverage of genes by exome sequencing. However, as shown in FIGURE 1B, greater average depth does not necessarily mean better sensitivity to detect variants, as some regions of a gene may have excessive coverage while others are completely missed. In addition, it is important to account for the variation in coverage for a particular gene from sample to sample. If this variance is great, there is a chance that a gene with good average coverage across the gene may still not be well covered in your patient's sample (FIGURE 1A).

A better measure with respect to accuracy is one which reflects the proportion of bases that are routinely sufficiently covered to call variants at that position, i.e. the sensitivity to detect variants in a particular gene. Disease-causing variants can simply be missed if the coverage for a particular gene is not sufficient across the whole gene, consistently, in every sample.

### Exome Sequencing Misses Non-coding Variants

By design, exome enrichments kits primarily target genomic regions that code for protein and so variants located in UTRs, intronic, promoter, and intergenic regulatory regions are missed. Although it is often difficult to interpret novel variants in such regions, there are known pathogenic variants in many genes that lie outside the exons. Such disease-causing variants cannot be detected by standard exome sequencing platforms.

### Exome Sequencing Misses Structural Variants

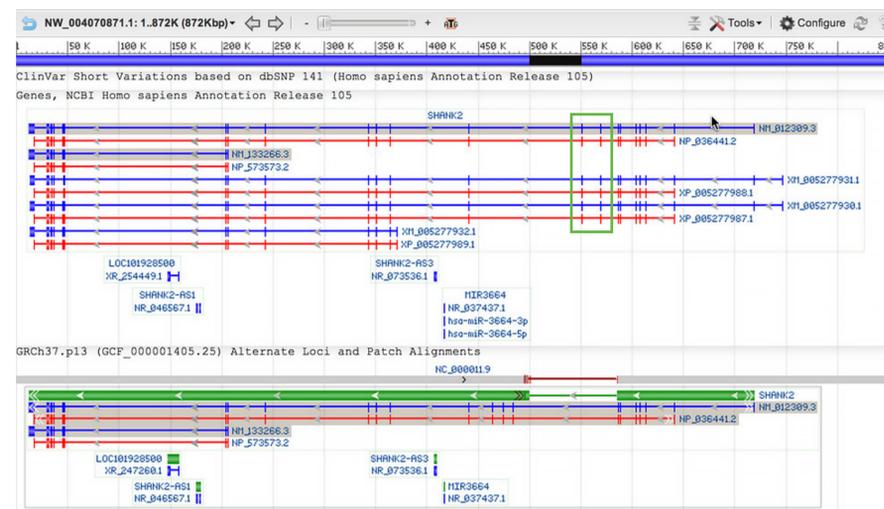
Tools for detecting structural variants have largely been developed for genome sequences and cannot be readily applied to exomes. Therefore, most exome sequencing services limit variant identification to SNVs and small indels thereby incompletely interrogating genes for mutations, leading to false negative results.

## Limitations Due to the Human Reference Genome Sequence

### Missing Genes in the Reference (false negative results)

The recognition of variants in a sample relies on sequence reads being aligned to the reference genome sequence, therefore robust gene representation in the reference assembly is critical to analyzing whole exome data. However, there are regions of the genome, including whole genes, that are missing from the current reference assembly, GRCh37. There are over 650 medically important genes that are impacted by the assembly update. In many cases (FIGURE 2) these updates add coding sequence to GRCh38 that is absent from GRCh37. Use of an incomplete reference sequence for alignment means that variants in such regions are not detected through exome sequencing.

FIGURE 2: Alignment of *SHANK2* to current reference GRCh37 and a "fix patch" in GRCh38



*SHANK2* is a gene associated with autism spectrum disorders for which clinical testing is currently available. Two exons of this gene are not included in the current human reference sequence, GRCh37. This means that sequence reads from this region are not included in exome alignments and therefore variants in these exons are not detected through exome sequencing.

### Missing Paralogs in the Reference (false positive results)

Genomic duplication that has occurred only in the human lineage gives rise to human specific sequences. These sequences show a high degree of sequence identity between paralogs and are difficult to sequence and assemble. Some such paralogs are therefore missing from the current genome reference assembly. However, many of these sequences have important biological functions and some have been associated with disease. Missing paralogs in the reference sequence may lead to false positive variant identification, as sequence reads from the missing paralog are misaligned to the paralog present in the reference. As an example, the *HYDIN* locus at 16q22 is associated with primary ciliary dyskinesia. There has been a human specific partial duplication of this gene, with the paralog residing at 1q21. The chromosome 1 and chromosome 16 loci share 99.4% identity over 300 Kb. The *HYDIN2* paralog was included in GRCh37 only as an unlocalized scaffold but has been added to the chromosome 1 assembly in GRCh38. All variation *HYDIN*, including clinically associated variants, needs to be reviewed in light of this highly related paralog. Variants in either region will be difficult to identify reliably with current sequencing technologies.

## Conclusion

In addition to the technical issues associated with variant detection outlined above, variant classification relies on a robust understanding both of the genetic variation that exists in the population as a whole as well as the phenotypic spectra of diseases, both of which are rapidly evolving. It is important that a genetic counselor engaged in ordering exome sequencing for their patients understands the various issues that can hamper a diagnosis being made by this approach. Familiarity with such issues enables informed test selection decisions, proactive discussions with laboratories, and appropriate follow-on patient care.

## Further Information

### Personalis @ NSGC 2014

- Sarah Garcia et al., "Use of an Enhanced Exome with Genome-wide Structural Variant Detection for the Diagnosis of Mendelian Disease" Poster #124
- Jeanie Tirch et al., "Successful Utilization of Enhanced Exome Sequencing to Identify the Genetic Cause of Retinal Disorders in a Case Series" Poster #148
- Sarah Garcia et al., "Homocystinuria Diagnosed by Whole Exome Sequencing in Siblings from an Isolated Central American Village" Poster #123
- Sarah Garcia & Gemma Chandratillake, "Application of an Enhanced Exome in Diagnosis of Rare Genetic Diseases" CEU Session @ Great Hall B&C, Friday 7:00 AM
- Gemma Chandratillake, "Revised Diagnosis Through Exome Sequencing of an Infant With Congenital Cataracts Expands Phenotypic Spectrum of COL4A1-associated Disorders" Concurrent Session: Genetic Testing I, Friday 8:15 AM

Booth  
205

