May 13th, 11:30 AM - 1:30 PM

A discrepancy between the human reference genome (GRCh37) and transcriptome (RefSeq) results in the incorrect annotation of a clinically-relevant sequence variant in RECQL4

Lisa A. Lansdon
Children's Mercy Kansas City

A discrepancy between the human reference genome (GRCh37) and transcriptome (RefSeq) results in the incorrect annotation of a clinically-relevant sequence variant in *RECQL4*

Submitting/Presenting Author (must be a trainee): Lisa A. Lansdon
Primary Email Address: lalansdon@cmh.edu

☐Resident/Psychology Intern
X Fellow

Primary Mentor (one name only): Midhat Farooqi
Other authors/contributors involved in project: Byunggil Yoo, Neil Miller, Maxime Cadieux-Dion, Emily Farrow, Elena Repnikova, Carol J. Saunders, Isabelle Thiffault, Midhat S. Farooqi

IRB Number: N/A

Describe role of Submitting/Presenting Trainee in this project (limit 150 words):

Dr. Lansdon is working to identify additional patients who harbor clinically relevant variants which are missed using standard informatics pipelines due to discrepancies between the human reference genome (GRCh37) and the RefSeq transcript sequence. In addition, she is helping to generate a catalog of these sequence discrepancies between GRCh37 and transcripts in RefSeq which can be employed in GRCh37 bioinformatic pipelines.

Background, Objectives/Goal, Methods/Design, Results, Conclusions limited to 500 words

Background:

A 17-month-old female was seen in genetics consultation for short stature, developmental delay, and a family history of a 10q26 duplication of unknown significance. She exhibited slight trigonocephaly, muscle stiffness, asymmetric kidneys, a small head circumference, bilateral epicanthal folds, microretrognathia, a high arched palate, and mild 2,3 syndactyly. Chromosomal microarray, thyroid function studies, 7-DHC, metabolic labs (acylcarnitine profile, plasma amino acids, and urine organic acids), telomere length studies, and chromosome breakage studies were non-informative. Overall, her phenotypic features were not suggestive of a specific diagnosis, though Floating-Harbor and Cornelia de Lange syndromes were suspected. Custom sequencing and a deletion/duplication panel of genes implicated in these disorders were negative for any clinically significant variants. A symptom-driven exome identified six variants of unknown significance, including *RECQL4*. We were particularly intrigued by *RECQL4* because pathogenic variants have been identified in individuals with Rothmund-Thomson, RAPADILINO, and Baller-Gerold syndromes, and we noted there was substantial phenotypic overlap between these syndromes and our patient. Our bioinformatics pipeline, based on human reference genome build GRCh37 (hg19) and

annotated using the annotation provided by RefSeq (version 1-13-2017), reported the variant in *RECQL4* (NM_004260.3) as c.3524C>T (p.Pro1175Leu). However, examination of this position in IGV showed that the reference amino acid in the transcript above was a glutamine. Based on the RefSeq transcript, a c.3523C>T variant would lead to the following protein change: (p.Gln1175*).

Objectives/Goal:

Determine the cause of the discordance between IGV and RefSeq.

Methods/Design:

By comparing the human reference genome sequence and RefSeq transcript sequence of *RECQL4*, we found a discrepancy between the two in exon 15. Specifically, the human reference genome carried an insertion of a cysteine residue at Chr8:g.145738769 (c.2296). This then resulted in a "frameshift" for all subsequent amino acid positions downstream with regards to variant effect prediction by our pipeline and accounted for the generation of a proline residue instead of glutamine at p.1175.

Results:

Our case reveals an instance of a potentially pathogenic variant (c.2296dup) in a clinically-relevant gene (*RECQL4*) being harbored by the human reference genome (GRCh37).

Conclusions:

Discrepant variants in GRCh37 reduce the sensitivity of variant calling and cause incorrect variant effect prediction since the latter is typically based off of the reference genome sequence. In addition, these insertions and deletions result in false variants called in databases, such as gnomAD, and standard informatics pipelines. Although use of GRCh38 resolves the *RECQL4* sequence discrepancy observed between GRCh37 and the RefSeq transcript (NM_004260.3)—due to removal of the inserted 'C' nucleotide in GRCh38—clinical implementation of this would require migration to GRCh38 and revalidation of the entire pipeline. Moreover, sequence discrepancies between GRCh38 and RefSeq also exist.

We reported that our patient harbored a likely pathogenic variant in *RECQL4,* c.3523C>T (p.Gln1175*) and was at least a carrier for a *RECQL4*-related disorder. We also recommended that deletion/duplication testing be performed for this gene. At minimum, we recommend that laboratories catalog genes which have a discrepancy between the human reference genome and transcriptome to aid in the recognition of this problem during analysis.