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

Identification of Clinically-Relevant Sequence Variants Within the Human Reference Genome

Lisa A. Lansdon
Children's Mercy Kansas City, lalansdon@cmh.edu

Identification of Clinically-Relevant Sequence Variants Within the Human Reference Genome

Lisa A. Lansdon, Maxime Cadieux-Dion, Byunggil Yoo, Neil Miller, Emily Farrow, Joseph T. Alaimo, Elena Repnikova, Isabelle Thiffault, Carol J. Saunders, Midhat S. Farooqi

Department of Pathology & Laboratory Medicine, Children's Mercy Kansas City, Kansas City, MO

Background and Objective

Background

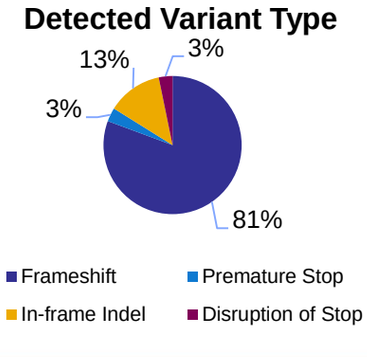
- Clinical exome and genome sequencing (CEGS) at our institution utilizes assembly GRCh37 of the Human Reference Genome for variant calling and annotation. However, GRCh37 contains known disease-associated variants such as Factor V Leiden (FVL): c.1601G>A (p.Arg534Gln). Individuals homozygous for this nonsynonymous change would not be detected by CEGS pipelines using GRCh37 as such individuals would not be variant from the reference genome. Although the existence of these disease-associated variants in GRCh37 is known in general, little work has been done to catalog their locations or determine their impact on clinical variant detection and interpretation.

Objective

- Identify potential disease-associated variants in GRCh37 and assess their effect on clinical variant interpretation.

Methods: Identification of Sequence Discrepancies

- We hypothesized that potential disease-associated variants carried by GRCh37 would occur at a high minor allele frequency (MAF) in CEGS databases. By filtering the Broad's gnomAD database and our internal CEGS variant database for variants occurring at MAF of 95% or greater, we systematically identified likely disease-associated sequences within GRCh37. We prioritized variants that resulted in a frameshift, insertion or deletion (indel), premature stop codon, or stop codon disruption.

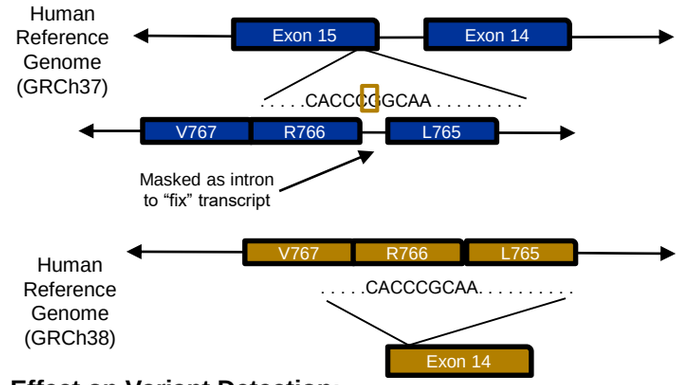


Results: Discrepant Regions and Variant Interpretation

- We identified 31 variants in GRCh37 within 28 genes associated with clinical phenotypes in OMIM.
- Comparison of GRCh37 with GRCh38 revealed that 24 out of the 31 variants (77.4%) identified in our study were resolved in the updated genome assembly.

Results: Variant Example – *RecQ-like helicase 4 (RECQL4)*

Call in gnomAD/CEGS DB: *RECQL4*:c.2297del (p.P766Rfs*444)
Variant in GRCh37: *RECQL4* (NM_004260.3):c.2296dup (p.V766Gfs*43)



Effect on Variant Detection:

- GRCh37 has an inserted G at chr8:145738768.
- To resolve this discrepancy, the insertion is masked in the RefSeq transcript. This generates an additional one base pair intron and splits exon 14 into two (exon 14 and exon 15).
- Patients harboring (NM_006331.7):c.2296dup will go undetected if using GRCh37.
- Sequence variants downstream of c.2296 (NM_004260.3) will have incorrect variant effect prediction if aligned to GRCh37.

Summary and Discussion

- Using population frequency data, we identified 31 variants within clinically-significant genes in GRCh37. These variants complicate variant detection in clinical exome and genome sequencing (CEGS) pipelines.
- The majority of variants identified are expected to result in a shift in the reading frame. These not only have implications for correct variant detection and effect prediction at the variant site but also all variants occurring downstream.
- Increased awareness and in-house annotation of disease-associated variants harbored by GRCh37 is necessary to ensure more accurate variant detection.

Future Directions

- Use the American College of Medical Genetics (ACMG) scoring criteria to assess the pathogenicity of each variant identified.
- Publish a reference list of discrepant variants in GRCh37.
- Use the reference list of pathogenic discrepant variants in GRCh37 to develop detection methods for homozygous variants at these loci.
- Evaluate our internal CEGS database to determine the number of carriers of pathogenic or likely-pathogenic discrepant variants.
- Assess feasibility of migration to GRCh38 for CEGS analysis at Children's Mercy Kansas City.

Children's Mercy KANSAS CITY

UMKC School of Medicine