Evaluation Of Data Quality In Public Databases Is Critical Before Its Use As Evidence In Clinical Variant Classification

Hannah C. Cox, Melanie Jones, Lisa Esterling, Debora Mancini-DiNardo, Thaddeus Judkins, Susan Manley, Karla R. Bowles, Bradford Coffee

Myriad Genetic Laboratories, Inc., Salt Lake City, UT

BACKGROUND

- Presence of a variant at a frequency greater than expected for the disorder in publically available research-grade population databases, such as the Genome Aggregation Database (gnomAD), is cited as strong evidence in ACMG/AMP guidelines that a variant is benign.[1,2]
- Conversely, the absence of a variant from population databases can be used as moderate evidence that it is pathogenic.[1]
- However, data taken without evaluation of its quality can lead to the misclassification of variants in the clinical laboratory.

OBJECTIVE

- Demonstrate that comprehensive assessment of data quality in public databases is needed to avoid variant misclassification.

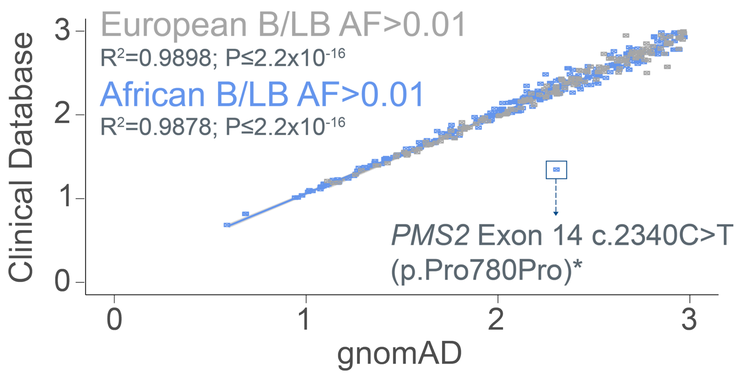
METHODS

- We evaluated variants identified during clinical testing by a single laboratory across 24 cancer-predisposition genes that were also present in the gnomAD exome data (release 2.0.2) (N=15,184).
- The analysis was restricted to the protein coding exonic regions and intronic regions up to -20 and +10 nucleotides.
- Quality parameters were assessed including gnomAD quality filters, coverage, variant root mean square mapping quality, and next-generation sequencing (NGS) read frequencies.
- The impact of clonal hematopoiesis, pseudogene homology, underlying sequence complexity, and ancestry were also assessed.
- Variants identified by the clinical testing laboratory were also compared to gnomAD with regards to whether the variant was present or absent in the public database.

RESULTS

- There was a strong correlation between allele frequencies (AF) for benign/likely benign (B/LB) variants (Figure 1).
- 7% (1,040/15,184) of shared variants did not pass the gnomAD quality filters.
- Decreased mapping quality (MQ) is an indication of pseudogene interference (Figure 2).
- 13% (1,895/14,144) of variants that passed gnomAD filters did not pass additional quality parameters (Figure 3), or showed evidence of somatic variant contamination (Figure 4).
- 75% (43,641/58,525) of variants identified by our laboratory were absent in gnomAD, including 9,882 (23%) variants classified as B/LB based on other evidence.

Figure 1. Correlation of Benign/Likely Benign Variant Allele Frequencies Between Databases



*This variant is in a region homologous to *PMS2CL*.

Figure 2. Pseudogene Interference Reflective in Public Database Mapping Quality

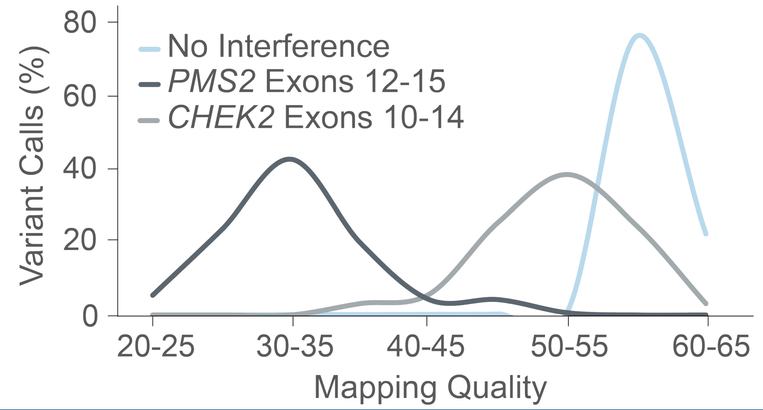
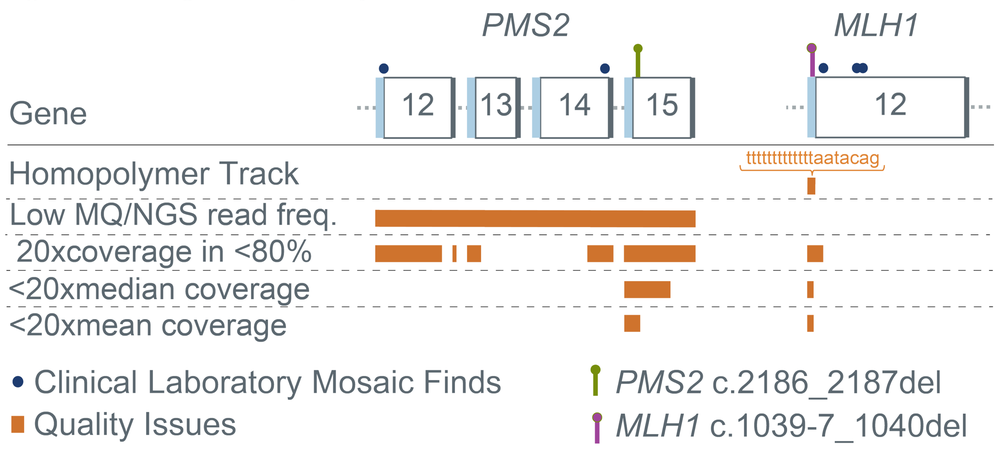
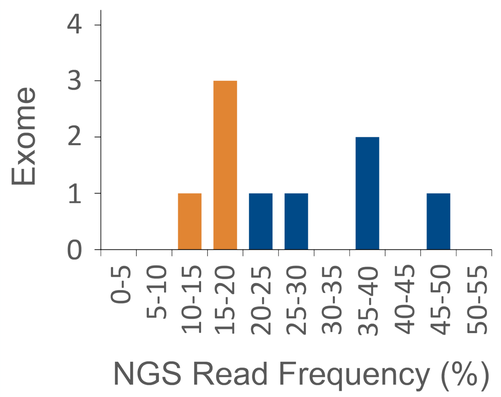


Figure 3. Regional Quality Issues in Public Database



PMS2 c.2186_2187del has an allele frequency of 2.57% in gnomAD African exomes, but has been observed only <.001% in clinical database.
MLH1 c.1039-7_1040del has an allele frequency of 0.15% in gnomAD Latino exomes, but has never been observed in the clinical database.

Figure 4. TP53 c.743G>A (p.Arg248Gln)*



*This is a known somatic hotspot with read frequencies of 10 - 50% in gnomAD.

CONCLUSIONS

- Though there is strong correlation in B/LB allele frequencies between our laboratory and gnomAD, confounding factors such as pseudogene interference, sequence complexity and somatically acquired variant contamination can cause allele frequencies to deviate significantly.
- Among variants detected in clinical testing that were absent in gnomAD, 23% (9,882/43,641) were classified as B/LB based on other evidence. Therefore, the use of absence of a variant in a general population database as moderate evidence of pathogenicity as ACMG/AMP guidelines suggest could potentially lead to variant misclassification.
- While all clinical testing laboratories utilize ACMG/AMP guidelines, these data highlight the importance of robust laboratory processes to ensure that only high quality data is used in variant classification.

REFERENCES

1. Richards S, et al. *Genet Med*. 2015:17(5);405-24.
2. Lek M, et al. *Nature*. 2016:536(7616):285-91.

Presented at ACMG on April 12, 2018.