Introduction

- Genomic screening of the general population for preventable, monogenic disease has potential to decrease morbidity and mortality.
- The selection of which genetic variants to return has tremendous impact on the specificity and positive predictive value of the test, which in turn has important downstream consequences for the success of any such endeavor.
- We selected 17 genes for 11 conditions that are among the most medically actionable of the Mendelian disorders for genomic screening.
- We screened 478 exome sequences for potentially pathogenic variants in these genes with 5 variant selection algorithms, and show the false positive rate of these algorithms.

Methods

- Variants from 478 exomes from a diagnostic sequencing study (NCGENES) were loaded into a PostgreSQL database (v.9.0.3) for annotation and facilitation of queries.
- Population allele frequency estimates were determined using the Exome Aggregation Consortium (ExAC), a resource composed of 63,358 unrelated individuals sequenced through a variety of studies.

The specificity, false positive rate, and number of variants returned per 1000 people screened was calculated for each of five variant selection algorithms:

- VSA-1 includes rare variants classified as “Pathogenic” in ClinVar. This is the least sensitive algorithm.
- VSA-2 adds rare predicted truncating variants (nonsense, frameshift, canonical splice-site).
- VSA-3 adds variants classified as “Likely Pathogenic” in ClinVar and/or as a “Disease Mutation (DM)” in HGMD.
- VSA-4 adds rare missense variants with CADD scores >13 that are located within a conserved functional domain.
- VSA-5 adds all rare missense variants, regardless of CADD score or location. This is the most sensitive algorithm.

The medical literature was reviewed in order to estimate the clinical sensitivity of diagnostic testing (corrected for locus heterogeneity), and the NNS based on the minimal sensitivity.

Results

- The NPV (A) and PPV (B) of genetic screening is demonstrated for prevalence values ranging across four log scales (from 10% to 0.01% prevalence).
- For any rare disease, the NPV is less influenced by the test characteristics while PPV has extreme dependence on specificity.

- The yield of potentially pathogenic variants using 5 variant selection algorithms with varying sensitivity is shown below. The number of people who would screen positive per 1000 individuals screened is displayed on the vertical axis.

Conclusions

- To optimize public health benefits from screening the genome for preventable, rare disease, the highest specificity must be ensured.
- Disease-specific false positive rates can be chosen, and different variant selection algorithms may be pursued depending on the presence of a confirmatory test and other downstream consequences of screening.

This work was supported by NHGRI 2P50 HG004488 (“Genescreen”), NHGRI 1U01HG006487-01 (“NCGENES”), and the Howard Holderness Distinguished Medical Scholars Foundation.