

MULTIGENIC ANALYSIS OF CONSUMER GENOMES FOR HEALTH RISK PREDICTION

M. Swan*

*Personal Genome Research, DIYgenomics, P.O. Box 61258
Palo Alto, CA 94306, USA*

**Email: Melanie@DIYgenomics.org*

Risk interpretation for common disease conditions may vary across direct-to-consumer (DTC) genomic services due to differences in three factors: the SNPs and loci selected for analysis, the average lifetime risk values assigned to the same underlying populations, and the quantitative risk assignment methodologies used. Simply because results differ does not mean that consumer genomic services are not useful since any variant with a polymorphism may indicate a higher risk for that condition. Multigenic risk assessment is a nascent field, and over time the sophistication of techniques applied by research scientists and consumer genomic companies could likely improve. Greater consistency in risk prediction is needed for the long-term validity, utility, and credibility of personal genomics.

1. INTRODUCTION

Direct-to-consumer (DTC) genomic services have become available in the last few years and have the potential for quick and substantial adoption as the cost of genomic sequencing continues to drop, the validity and utility of genetic data in chronic disease prevention increases, individuals are motivated to investigate their own genomic data, and physicians begin to incorporate genomics into care regimens.

There are three types of direct-to-consumer genomic services currently available to individuals without a physician's prescription: one-off SNP (single nucleotide polymorphism) tests for specific conditions and paternity tests, multi-SNP risk assessment tests mapping several SNPs to dozens of disease conditions, and whole human genome sequencing assessing hundreds of disease risks.

Gene carrier status and pharmacogenomic data may be detectable from single SNPs, but disease prediction is more challenging. Research concerning multigenic common disease conditions such as diabetes, cardiovascular disease, and cancer is new in the last several years. The potentially hundreds of SNPs on multiple genes that may relate to conditions have not yet been understood definitively from a scientific perspective. The accuracy of raw genotyping data has been found to be consistent across DTC services, but risk interpretations for multigenic conditions may vary considerably.

2. VARIANCE IN INTERPRETATION

Multigenic condition risk interpretation may vary between DTC genomic services due to differences in three factors: the SNPs and loci selected for analysis, the average lifetime risk values assigned to the same underlying populations, and the quantitative risk assignment methodologies used.

The first reason that DTC genomic companies have different risk assessments for the same condition is that they are evaluating different SNPs. Each company has rigorous criteria for selecting underlying research studies and SNPs, but the methodologies vary and have discretionary components. DTC genomic companies are quick to point out that they are aware that different companies look at different SNPs and attribute this to the use of surrogate SNPs at the same locus. However, not only are SNPs different, but also loci (less than 20% of loci are evaluated by more than two companies¹), so selection criteria is the main reason that SNPs and loci vary.

Just as DTC genomic companies may choose different studies to select SNPs, they may choose different studies to obtain lifetime risk averages for the underlying populations. There are considerable differences in the supporting epidemiologic literature due to the general challenge of predicting lifetime risk for conditions. Widely agreed upon figures do not exist for most conditions, and also vary by age tiers. There is significant variance in the average lifetime risk for

* Corresponding author.

underlying populations cited by DTC genomic companies for certain conditions.

A third factor leading to variance in overall DTC genomic company risk assessment is different calculations of quantitative risk values, often for the same genotype and cited research study. This can cause risk assessment to vary in magnitude and sometimes directionally. In addition, a multiplicative technique is used to derive the composite risk for each condition (e.g.; the product of the odds ratios or other quantitative metric for all SNPs is taken). A multiplicative technique is problematic in that it does not allocate more weight to strong-effect SNPs, does not take into account protective SNPs, and could change over time as more SNPs are found to be associated with conditions.

3. CONCLUSION

An analysis of multigenic condition risk assessment across DTC genomic services from publicly-available materials reveals three reasons that risk interpretation may differ. First, different SNPs and loci are evaluated. Second, different underlying lifetime risk averages for general populations are employed. Third, different methodologies are used in the assignment of quantitative risk values. Further, some of the most critical SNPs may not be evaluated due to patents or other issues (for example, ApoE4 and BRCA1/2 are not evaluated; other SNPs are reviewed for Alzheimer's disease and breast cancer).

Simply because there is diversity in DTC multigenic risk assessment does not mean that the services are useless since any variant with a polymorphism may indicate a higher risk for that condition. Genomic information is actionable now in routing higher-risk individuals to earlier screenings and in categorizing drug responders and non-responders. Going forward, the validity, utility, and credibility of DTC personalized genomic services may be improved by standardizing the core variants that are reviewed for conditions, identifying appropriate lifetime risk averages, and establishing accurate risk assessment methodologies for multigenic conditions.

References

1. Swan M. Multigenic condition risk assessment in direct-to-consumer genomic services. *Genet. Med.* 2010; **May**;12(5):279-88.