Personalized Medicine Coalition

Personal Genomics and Industry Standards: Scientific Validity

Participants: 23andMe, deCODE, Navigenics.

Background

In July of 2008, three personalized genomics companies agreed to work with the Personalized Medicine Coalition on a set of standards regarding the scientific validity of gene scans available to consumers directly, over the internet. The following summary describes the process.

Analytical Validation

Genotyping Accuracy

The companies use a single nucleotide polymorphism (SNP) genotyping platform in a federally regulated laboratory. The DNA chips used covering 600,000 to 1millions SNP markers have 4 to 20 fold redundancy for each SNP measured resulting in better than 99% accuracy**.** For example, the Illumina arrays run by deCODE and 23andme exceed 99.9% accuracy when compared to other genotyping platforms or bidirectional sequencing).

Risk calculation methods

All companies provide the individual's estimate of disease risk in both absolute lifetime risk and relative risk compared to the general population, so that physicians are able to use the relative genetic risk in conjunction with relative risk for other risk factors. One reason for giving the estimated lifetime risk to the individual is to emphasize that there is a large range of baseline lifetime risks across common diseases.

All three companies use standard statistical methodologies to convert from risk numbers reported in peer-reviewed published studies (allelic or genotypic odds ratio, or OR) to risk compared to the general population as well as lifetime risks or lifetime incidences. All three companies generally apply a multiplicative model when combining multiple markers together. This is justified because no other model fits the data better based on the large datasets available for common variants conferring risk to common disease. The companies will make exceptions if published scientific data support it.

The two key differences between the companies is the timing of conversion for OR to risk compared to the general population and the assumptions made about controls. deCODE converts allelic OR for each marker to risk compared to the general population before combining marker risks (a risk multiplicative model). In this model, allelic OR is

assumed to be equal to the relative risk between heterozygotes versus noncarriers, and between homozygotes of risk allele versus heterozygotes. deCODE also generally assumes that the controls used in the literature used to define the allelic OR are population-based controls rather then unaffected controls — however, if the controls are clearly unaffected controls beyond the peak age of incidence for the disease, then unaffected controls are assumed.

23andme and Navigenics generally assume unaffected controls were used in the estimating allelic or genotype OR in most studies. Therefore, allelic OR is a better estimate of disease OR. 23andme combines markers using a multiplicative model (multiplicative odds ratio model) before converting to risk compared to the general population (taking into account the disease prevalence). Navigenics, like deCODE, converts each marker to relative risk compared to the general population (but needs to take into account disease prevalence for conversion) and then uses the same multiplicative risk model as deCODE to combine the markers. Therefore, Navigenics is using a model that is a hybrid between deCODE's and 23andme's approaches. The estimated risk derived from all three models were expected to generally agree for over 99% of individuals. It was also expected that at the extremes of high risk there would be an overestimate of risk using deCODE's approach and an underestimate of risk using the other two approaches, mainly driven by the different assumptions about population controls versus unaffected controls in the literature.

All three models were run against three diseases using the same assumptions: same markers, allelic frequencies, allelic OR, and disease lifetime risk. The models were run against the Caucasians of European origin in Utah, or CEU set, of HapMap samples as a source of genotyped individuals. All three methods agreed exactly in terms of ranking of individuals in terms of their risk. As expected, the lifetime risks agreed across all three methods within five percent absolute lifetime risk with the exception of the rare extremes. For example, modeling the risk from eight widely replicated breast cancer variants, the estimated lifetime risks agreed within 5% (absolute difference) of each other for 99.7% of the population. At the upper 0.3% percentile of risk, deCODE's estimates ranged from 32.7% to 56.4% (weighted average risk of about 35% and 23andme's estimates ranged from 27.7% to 39.8% (average risk of about 29%) estimated lifetime risk for breast cancer, both assuming average lifetime risk of 13% for white females. Navigenics was within these estimates.

Clinical Validation

Epidemiological sources

All three companies choose one or more epidemiologic studies published in the

literature to define or derive lifetime risk or cumulative incidence. In July of 2008 the companies noticed that the assumed baseline lifetime risk was the greatest source of differences in lifetime risks reported out by the three companies to the same individual. Therefore, all three companies are working together to use the same average lifetime risks based on epidemiologic studies where possible if the phenotype definition used is the same. Some of the differences stem from a different phenotype definition (for example, ruptured abdominal aneurysms versus ruptured plus asymptomatic aneurysms, or celiac disease based on cases diagnosed through the health care system versus those found through screening the general population). Others come from quoting different published studies defining lifetime risk. Some of the differences come from 23andme using cumulative incidence and deCODE and Navigenics using lifetime risk. The companies are working to come up with consensus numbers to use where possible. Any discrepancies in lifetime risk and cumulative incidence scores will be explained transparently on our respective websites and in our consumer guide.

SNPs used for genetic risk

All three leading companies agreed to use only SNPs that are clinically validated — that is, replicated in at least two well-powered studies. 23andme separates its clinically validated reports from “research reports”, which are not necessarily clinically validated. Most of the SNPs currently used by the three companies have been replicated in several case–control studies totaling thousands of cases and controls. The companies use the allelic OR or genotypic OR derived from large studies. There is no requirement that the same SNPs or the same number of SNPs be used by the three companies — however, the SNPs selected should be clinically validated.

Transparency

Transparency in disclosing companies’ methods and criteria is the most pragmatic goal, instead of achieving identical risk estimates among the three services. These three companies are committed to the current practice of disclosing their respective risk analysis methodologies on their website. However, as described above, the genetic risk profiles agree very well across the three companies when a similar number of validated SNP markers are used.

Consumer and health care provider guide

Another possible contribution to transparency discussed by industry leaders is to add examples of similarities and differences in the three companies' methodologies in a PMC-developed guide to these products. This information could easily be drawn from

the companies' glossaries and sample calculations, which are all readily available on their websites.

Methodology links

23andMe: https://www.23andmeobjects.com/res/3296/pdf/2301_Estimating_Genotype_Specific_Incidence.pdf

DeCode: www.decode.com/standards

Navigenics:http://www.navigenics.com/static/pdf/Navigenics%20White%20Paper.pdf

Personalized Medicine Coalition

The Personalized Medicine Coalition, representing a board spectrum of academic, industrial, patient, provider and payer communities, seeks to advance the understanding and adoption of personalized medicine concepts and products for the benefit of patients. More information can be found at www.PersonalizedMedicineCoalition.org.