Variations in predicted risks in personal genome testing for common complex diseases

Rachel R.J. Kalf, MSc[1], **Raluca Mihaescu, MD**[1], **Suman Kundu, MSc**[1], **Peter de Knijff, PhD**[2], **Robert C. Green, MD, MPH**[3], and **A. Cecile J.W. Janssens, PhD**[1,4]

[1]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands [2]Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands [3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA [4]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Abstract

Purpose—The promise of personalized genomics for common complex diseases depends, in part, on the ability to predict genetic risks on the basis of single nucleotide polymorphisms. We examined and compared the methods of three companies (23andMe, deCODEme, and Navigenics) that have offered direct-to-consumer personal genome testing.

Methods—We simulated genotype data for 100,000 individuals on the basis of published genotype frequencies and predicted disease risks using the methods of the companies. Predictive ability for six diseases was assessed by the AUC.

Results—AUC values differed among the diseases and among the companies. The highest values of the AUC were observed for age related macular degeneration, celiac disease, and Crohn disease. The largest difference among the companies was found for celiac disease: the AUC was 0.73 for 23andMe and 0.82 for deCODEme. Predicted risks differed substantially among the companies as a result of differences in the sets of single nucleotide polymorphisms selected and the average population risks selected by the companies, and in the formulas used for the calculation of risks.

Conclusion—Future efforts to design predictive models for the genomics of common complex diseases may benefit from understanding the strengths and limitations of the predictive algorithms designed by these early companies.

INTRODUCTION

It is envisioned that genome testing will personalize medicine, not only for the diagnosis and treatment of monogenic or Mendelian disorders, but also for the prevention of common complex diseases such as type 2 diabetes, age-related macular degeneration, and heart attack. Since 2007, personal genome tests have been offered directly to consumers via the Internet to educate and empower consumers about the risk of common diseases.[1–4]

Common complex diseases are caused by an interplay between multiple genetic and non-genetic factors.[5] Genomewide association studies are rapidly discovering variants implicated

in common disease but to date still leave a large part of the heritability unexplained because the identified single nucleotide polymorphisms (SNPs) generally have minor effects on disease risk.[6] Consequently, genetic risk models based on known SNPs typically have a low to moderate predictive ability for most diseases. Exceptions do occur when one or more variants have a strong effect on disease risk, as in age-related macular degeneration and type 1 diabetes.[5,7,8]

The predictive ability of direct-to-consumer personal genome tests has not been demonstrated in empirical studies. Insights concerning the concordance of personal genome tests conducted by different companies are available from a few reports of individuals who had sent their saliva to more than one company.[9,10] These reports showed that predicted risks differed among companies and were divergent for some traits in some individuals.[9,10] Differences in predicted risks were attributed to variations in the selection of the SNPs used, their effect sizes, and the average population risks of disease that were used to calculate disease risks.[9,11–13] As genotyping and sequencing become less expensive, they will be entering the medical mainstream. The methods used for estimating the predictive ability of common variants to generate risk information will be an important concern. In anticipation of this, we conducted an in depth analysis and comparison of the approaches of the companies that pioneered the predictive use of genotyping in order to better understand the strengths and limitations of the methods they used to compute estimates.

We assessed and compared predicted risks and the predictive ability of personal genome testing offered by three companies: 23andMe, deCODEme, and Navigenics. The study was conducted in a hypothetical population of 100,000 individuals. Predicted risks were calculated using the methods of the companies, which were obtained from their websites. The predictive ability of the genetic risk models was quantified by the area under the receiver operating characteristic curve (AUC).

MATERIALS AND METHODS

Predicted risks and the predictive ability of personal genome tests from 23andMe, deCODEme, and Navigenics were assessed for six diseases: age-related macular degeneration, atrial fibrillation, celiac disease, Crohn disease, prostate cancer, and type 2 diabetes, which for all companies constitute a subset of all diseases tested. These diseases were chosen because of differences in the effect sizes of the SNPs discovered to date and differences in average population risks. Age-related macular degeneration and celiac disease are influenced by a few SNPs with strong effects on disease risk, whereas the other diseases are influenced by many SNPs with relatively weak effects. Celiac disease and Crohn disease are rare disorders, whereas the others are more common.

Because there are no prospective empirical data on the predictive ability of personal genome tests, we used hypothetical data to answer our research questions. A detailed description of the construction of the data sets, the calculation of predicted risks, and our efforts to verify correct interpretation of the risk calculation methods is provided in the Supplementary Materials and Methods online.

Simulated data

Construction of genotype data—Simulated data sets were constructed using a modeling procedure that has been validated and described in more detail elsewhere.[14,15] In short, this procedure creates genotypes for a hypothetical population of 100,000 individuals. For each SNP, genotypes are assigned randomly to individuals in such a way that genotype or allele frequencies in the 100,000 individuals match prespecified input values (see Supplementary Materials and Methods online).

Calculation of predicted risks—Predicted risks were calculated using the methods of 23andMe, deCODEme, and Navigenics, which were described on their websites or in downloadable white papers.[16–18] To calculate disease risks, all three methods require information on the average "population risk" and on the odds ratios and genotype or allele frequencies of the SNPs included in the test. The average population risks and the SNPs were obtained from the websites of the companies, and the odds ratios of the SNPs were extracted from the scientific studies referenced on the websites (accessed January 2012).[1–3] Genotype and allele frequencies were obtained from HapMap release 24 for 23andMe, cited scientific studies for deCODEme, and the company's website for Navigenics. The companies first compute the likelihood ratio or relative risk for each SNP using the odds ratio and genotype or allele frequencies. To generate predicted risks, these likelihood ratios or relative risks are combined with the average population risk (see Supplementary Materials and Methods online). All risks were calculated for Caucasian men.

Data analysis

To compare predicted risks among the three companies, we constructed one large data set with genotypes for the 113 SNPs tested by the three companies for all six diseases on the basis of genotype frequencies from HapMap release 28. For each individual, predicted risks were obtained using the formulas of the three companies, which yielded 18 predicted risks (6 diseases × 3 companies) per person.

To assess and compare the predictive ability, we used the genotype frequencies that the companies each used for the calculation of the likelihood ratios or relative risks (see above). Hence, we constructed hypothetical populations for each company and each disease separately. The predictive ability was quantified by the AUC.[19] The AUC values range from 0.5 (random prediction) to 1.0 (perfect prediction). The AUC represents the probability that a random individual who will develop the disease has a higher predicted risk than a random individual who will not develop the disease. For the calculation of the AUC, disease status was randomly assigned to individuals on the basis of their predicted risks, in such a way that for individuals with the same disease risk, the percentage of individuals who will develop the disease equals that risk when the subgroup of individuals with that risk would have been sufficiently large.[14] In other words, the simulation method assumes perfect calibration of the prediction models. To illustrate the predictive ability, we obtained the distribution of predicted risks for people who will develop the disease and those who will not across the three risk categories that 23andMe distinguishes in the presentation of disease risks on the personal webpages of their consumers. The thresholds for these categories of decreased, typical, and elevated risk are 20% below and above the average population risks (relative risks 0.83 and 1.2).[1]

Finally, we assessed the agreement between the companies in classifying each individual to the same risk category. We used the original large data set, constructed for the comparison of predicted risks among the companies, to assess the agreement in classification across the three risk categories that 23andMe distinguishes. All analyses were performed using R version 2.12.1.[20]

RESULTS

Table 1 shows that 23andMe, deCODEme, and Navigenics used similar average population risks for the prediction of disease risks, except for age-related macular degeneration and celiac disease. For celiac disease, deCODEme used an average population risk that was eightfold higher than that used by 23andMe and 16-fold higher than that used by Navigenics. The number of SNPs that were used for the calculation of the risk varied substantially among the companies. For the calculation of type 2 diabetes risks, 23andMe

used 11 SNPs, deCODEme 21, and Navigenics 18; and for prostate cancer the companies used 12, 26, and 9 SNPs, respectively. For four diseases, deCODEme used the most SNPs, and for all six the company used twice as many SNPs as 23andMe used. The Supplementary Table S1 online shows that most SNPs tested by 23andMe or Navigenics were tested by two or more companies but that deCODEme tested many SNPs that were not covered by the other companies.

Table 2 shows that for each disease the AUC of the tests differed among the companies. The largest difference was observed for celiac disease (0.73 for 23andMe and 0.82 for deCODEme). The AUC values were also substantially different among the diseases. The AUC values were around 0.80 for age related macular degeneration, celiac disease, and Crohn disease, but only around 0.60 for atrial fibrillation, prostate cancer, and type 2 diabetes. Table 3 illustrates the predictive ability using the risk categories defined by 23andMe. When the AUC values are higher, individuals who will develop the disease more often have elevated risks and individuals who will not develop the disease more often have decreased risks of disease. When the AUC values are closer to 0.50, the distribution of predicted risks across the risk categories is more similar, which reflects that the risk model does not discriminate between the two groups.

Figure 1 and Supplementary Figure S1 online show comparisons of predicted risks from the three companies for individual consumers. The strongest agreement in predicted risks was observed for atrial fibrillation, for which 23andMe and Navigenics predicted similar risks based on the same SNPs, but many consumers received substantially different risk assessments from the companies for other diseases. For example, for Crohn disease, 23andMe used variants that had higher effect sizes than those used by Navigenics and variants that were not covered by deCODEme (see Supplementary Table S1); and for celiac disease, deCODEme predicted higher risks than 23andMe due to the higher average population risk that was used in the calculation (Table 1).

Figure 1 also shows that both deCODEme and Navigenics used formulas that allowed predicted risks to be >100%. The highest risks in our hypothetical population, 327% by deCODEme and 193% by Navigenics, were predicted for age-related macular degeneration. We examined the extent to which differences in the formulas could explain the prediction of risks >100% by applying the three formulas to the input data (average population risk, odds ratios, and allele frequencies of the SNPs) of 23andMe (see Supplementary Figure S2 online). Supplementary Figure S2 shows that, in the range of higher predicted risks, the formulas of deCODEme and Navigenics produced higher risks than those of 23andMe and that these risks could exceed 100%, as was shown for atrial fibrillation and prostate cancer.

Finally, again using the risk categories defined by 23andMe (see Materials and Methods section), we investigated the extent to which the three companies assigned individuals to the same risk category (Table 4). The highest concordance was observed for celiac disease, for which 89.0% of the individuals were assigned to the same risk category (75.3% as decreased risk and 13.8% as elevated risk), which is explained by the fact that all three companies test for the same variant that had a strong effect. For other diseases, concordance ranged from 33.6% (prostate cancer) to 68.0% (age-related macular degeneration). In most other instances, two companies assigned an individual to the same risk category and the third company predicted an average risk. Yet, for Crohn disease, age-related macular degeneration, and prostate cancer, 27.1%, 19.9%, and 15.5% of the individuals, respectively, were predicted opposing risks by at least two companies.

DISCUSSION

In 2008, 23andMe, deCODEme, and Navigenics, in collaboration with the Personalized Medicine Coalition, published a white paper in which they described the strategies they used for calculating genetic risks of disease.[21] The companies explained and acknowledged that they use different SNPs, average population risks, and formulas to obtain predicted risks for consumers. Our analyses show that these differences in the SNPs, average population risks, and formulas yield substantial differences among the companies in the predictive ability for each disease and in predicted risks for individual consumers.

Before commenting on our results, three methodological issues may require further elaboration. First, we used simulated data to investigate the predictive ability of personal genome tests, because evidently empirical data were not available. On the basis of published genotype frequencies, we constructed genotype data for a hypothetical population of 100,000 individuals under the assumption that genetic variants inherit independently. Although this simulation method assumes perfect calibration of the risk models, which theoretically might lead to overestimation of AUC, we recently showed that this modeling approach was able to accurately replicate the AUC values of empirical prediction studies.[15] We therefore believe it is reasonable to assume that the use of simulated data does not distort the results of this study. Second, we applied the risk categories utilized by 23andMe, which have relatively low thresholds to define risks as being decreased or increased. When individuals are easily classified in the very broad decreased or elevated risk categories, the agreement in assigning an individual to the same risk category, as presented in Table 4, is likely overestimated. And third, all companies in our study provide regular updates of risk predictions to consumers when new SNPs are discovered or when better epidemiological data are available.[22] We performed our analyses in January 2012 and verified all input data in December 2012. The most important change in that period was that Navigenics was acquired by Life Technologies and deCODEme by Amgen, and both no longer offer personal genome testing.[2,3] 23andMe had updated the prediction of age related macular degeneration by the addition of two SNPs.[1] Our results should therefore be interpreted as a historical comparison of direct-to-consumer personal genome testing and as an illustration of how differences in the sets of SNPs selected, the average population risks, and the formulas used for the calculation influence predicted risks and the predictive ability of personal genome tests.

The predictive ability of genetic tests as assessed by the AUC indicates the extent to which the test, at the "population" level, can discriminate between people who will develop the disease and those who will not. In contrast, a comparison of predicted risks indicates the extent to which "individual" consumers receive different predicted risks from the companies. Our study showed that the predictive ability differed among the companies for each of the diseases, and that differences in predicted risks were substantial even when tests had similar predictive ability. We also observed that, in exceptional cases, predicted risks of deCODEme and Navigenics could exceed 100%. We investigated three main factors that have an impact on predictive ability and predicted risks and that might explain these observations.

First, the companies included a different number of SNPs in their genetic risk models. For most diseases, the tests of deCODEme included the same SNPs as 23andMe and Navigenics, as well as additional SNPs that were not covered by the others. More SNPs generally implies more differentiation in predicted risks, as indicated by a higher AUC, and gives different risk predictions for individual consumers. For example, 23andMe and Navigenics predicted similar risks for atrial fibrillation (both AUC = 0.58) because both considered the same two SNPs, whereas deCODEme considered four additional SNPs that introduced more

variability in predicted risks and led to slightly higher predictive ability (AUC = 0.62; Table 2 and Figure 1). Note that tests with the same AUC do not necessarily predict the same risks at the individual level. Despite similar AUC values (0.61 and 0.60), 23andMe and Navigenics predicted markedly different risks for prostate cancer. In general, similar AUC values mean that the tests perform equally in identifying at-risk individuals at the population level, but individual consumers may be selected in the at-risk group on the basis of one test and not on the other when the predictive ability is not perfect, and the tests consider different risk factors. This may even occur when the AUC of the test is high, as was demonstrated for age related macular degeneration, for which 20% of the consumers received risks in opposite risk categories (Table 4). Therefore, even tests that have appreciable predictive ability at the population level may have contradictory results for individuals.

Second, all three companies used an estimate for the population disease risk as the starting point for their predictions. Some of these averages were relatively similar, but others were markedly different. For age-related macular degeneration, average risks were up to 2.5-fold higher, and for celiac disease up to 16-fold higher among the companies. Differences in average risks do not affect the predictive ability of the test, because they increase or decrease risks of the entire population to the same extent, but they do have an impact on actual values of predicted risks. This was most clearly demonstrated for celiac disease, for which almost all predicted risks by deCODEme were higher than those predicted by 23andMe and Navigenics, because their average population risk was up to 16-fold higher. The companies have likely used different epidemiological studies to obtain their estimates, but it is unlikely that differences in study population and design can explain the large differences in the average population that are used. It is more likely that some are prevalence and others are incidence estimates, or that the estimates are obtained from studies with different follow-up times, yielding different proxies for the lifetime risk. These inferences raise the question of whether the companies are calculating risks on the basis of information that is relevant to their consumers. Most genome-wide association studies are conducted in Caucasian populations, and the odds ratios from these studies may not be relevant for other ethnicities. Also, the companies used average estimates of lifetime risks and did not take age into account for the calculation of risk, but the remaining lifetime risks are not the same for 20- and 60-year-olds. And consumers might be more interested in short-term, e.g., 10-year, risks than lifetime risks, because these better reflect the risk of becoming ill at younger ages. A more in-depth reflection is needed on what risks are most appropriate to return in personal genome testing.

And third, the companies applied different formulas, which affected the exact prediction of risks. A difference among the formulas is that deCODEme multiplied the likelihood ratio of a genotype combination (genetic profile) by the average risk, Navigenics multiplied the relative risk by the lowest possible risk, where 23andMe multiplied the likelihood ratio by the average odds.[16–18] These approaches yield similar predictions for lower risks, but the formulas of deCODEme and Navigenics appear to overestimate risks when predicted risks are higher. This difference in the calculation also results in scenarios in which predicted risks might become >100% for deCODEme and Navigenics (Figure 1), an observation that was previously made in a study on breast cancer risk.[23] The strategy of 23 and Me follows the widely accepted Bayes' theorem, which is in line with logistic regression and which prevents the resulting risks from exceeding 100%. DeCODEme multiplied likelihood ratios by the average risk, which is only appropriate when risks are small (see Supplementary Figure S2 online). Finally, Navigenics multiplied relative risks by the lowest possible risk, a method that becomes computational infeasible on a standard computer for risk models that involve more than 14 SNPs. The question of which method is the most appropriate is difficult to answer, because it is unknown which model best reflects the underlying biological pathways to disease.[24] Choosing the most appropriate computational method may

improve calibration of risks, and potentially the predictive ability, but this improvement is likely minimal as compared with the improvement that could have been achieved if non-genetic risk factors were considered in the prediction of disease.

The differences in the selected SNPs, average disease risks, and formulas have different impacts on the predicted risks and the AUC values. They all determine the exact values of the predicted risks, but only the selected SNPs have an impact on AUC values. In general, the more SNPs included in the risk model and the higher their odds ratios and genotype frequencies, the higher the value of the AUC. Differences in average risks and in the formulas do not affect the AUC values because AUC is essentially a rank test, and these differences do not change the rank order of the predicted risks. The differences in allele frequencies and odds ratios, given that the companies used different sources to obtain this information, would seem to be a possible explanation for the observed differences in the AUC values. Yet AUC is known to be relatively insensitive and unable to detect minor improvements of risk models.[25] The differences in odds ratios and allele frequencies were likely too minor to cause variation in the AUCs. The differences in the AUCs among the companies are predominantly explained by the selection of the SNPs.

In the absence of prospective empirical data, our study provided insight into the methodology and performance of risk estimation for personal genome tests. We showed that the predictive ability of personal genome tests and the predicted risks for individual consumers differed among the companies due to the differences in the SNPs selected, the average population risks, and the formulas. For six diseases, we showed that the personal genome tests of the three companies had limited predictive ability (atrial fibrillation, type 2 diabetes, and prostate cancer), a considerable (20–27%) probability of receiving "opposite" predictions (age-related macular degeneration and Crohn disease), or substantial differences in absolute risks at the individual level (celiac disease). These observations on the variation and pitfalls in disease risk predictions by personal genome tests provide insights into models of risk estimation and will inform the evolving discussion about the best use of genomic information in the consumer marketplace and in the practice of medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. 23andMe. [Accessed 1 January 2012 and 1 December 2012] https://www.23andme.com/

2. deCODEme. [Accessed 1 January 2012 and 1 December 2012] http://www.decodeme.com/

3. Navigenics. [Accessed 1 January 2012 and 1 December 2012] http://www.navigenics.com/

4. Helgason A, Stefánsson K. The past, present, and future of direct-to-consumer genetic tests. Dialogues Clin Neurosci. 2010; 12:61–68. [PubMed: 20373667]

5. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet. 2008; 17(R2):R166–R173. [PubMed: 18852206]

6. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet. 2012; 90:7–24. [PubMed: 22243964]

7. Hecker LA, Edwards AO, Ryu E, et al. Genetic control of the alternative pathway of complement in humans and age-related macular degeneration. Hum Mol Genet. 2010; 19:209–215. [PubMed: 19825847]

8. Eleftherohorinou H, Wright V, Hoggart C, et al. Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. PLoS ONE. 2009; 4:e8068. [PubMed: 19956648]

9. United States Government Accountability Office. [Accessed 1 January 2012] Direct-to-consumer genetic tests: Misleading test results are further complicated by deceptive marketing and other questionable practices. 2010. www.gao.gov/assets/130/125079.pdf

10. Ng PC, Murray SS, Levy S, Venter JC. An agenda for personalized medicine. Nature. 2009; 461:724–726. [PubMed: 19812653]

11. Imai K, Kricka LJ, Fortina P. Concordance study of 3 direct-to-consumer genetic testing services. Clin Chem. 2011; 57:518–521. [PubMed: 21159896]

12. Swan M. Multigenic condition risk assessment in direct-to-consumer genomic services. Genet Med. 2010; 12:279–288. [PubMed: 20474084]

13. Kuehn BM. Risks and benefits of direct-to-consumer genetic testing remain unclear. JAMA. 2008; 300:1503–1505. [PubMed: 18827199]

14. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med. 2006; 8:395–400. [PubMed: 16845271]

15. Kundu S, Karssen LC, Janssens AC. Analytical and simulation methods for estimating the potential predictive ability of genetic profiling: a comparison of methods and results. Eur J Hum Genet. 2012; 20:1270–1274. [PubMed: 22643180]

16. Macpherson, M.; Naughton, B.; Hsu, A.; Mountain, J. [Accessed 1 July 2012] White paper 23-01: estimating genotype-specific incidence for one or several loci. 2007. https://23andme.https.internapcdn.net/res/pdf/HIC-SXIYiYqXreldAxO5yA_23-01_Estimating_Genotype_Specific_Incidence.pdf

17. deCODEme. [Accessed 1 July 2012] SNP selection and risk calculation. demo.decodeme.com/healthwatch-information/risk-calculation

18. Navigenics. [Accessed 1 July 2012] The science behind the Navigenics service. www.navigenics.com/static/pdf/Navigenics-TheScience.pdf

19. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]

20. Ihaka GR, Gentelman R. R: a language for data analysis and graphics. J Comput Graph Stat. 1996; 5:299–314.

21. The Personalized Medicine Coalition. [Accessed 1 December 2012] Personal genomics and industry standards: scientific validity. http://www.personalizedmedicinecoalition.org/sites/all/themes/zen_pmc/documents/PMC%20personalgenomicsSci%20Valid15dec08.pdf

22. Mihaescu R, van Hoek M, Sijbrands EJ, et al. Evaluation of risk prediction updates from commercial genome-wide scans. Genet Med. 2009; 11:588–594. [PubMed: 19636253]

23. So HC, Sham PC. Effect size measures in genetic association studies and age conditional risk prediction. Hum Hered. 2010; 70:205–218. [PubMed: 20838054]

24. Moonesinghe R, Khoury MJ, Liu T, Janssens AC. Discriminative accuracy of genomic profiling comparing multiplicative and additive risk models. Eur J Hum Genet. 2011; 19:180–185. [PubMed: 21081969]

25. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation. 2007; 115:928–935. [PubMed: 17309939]
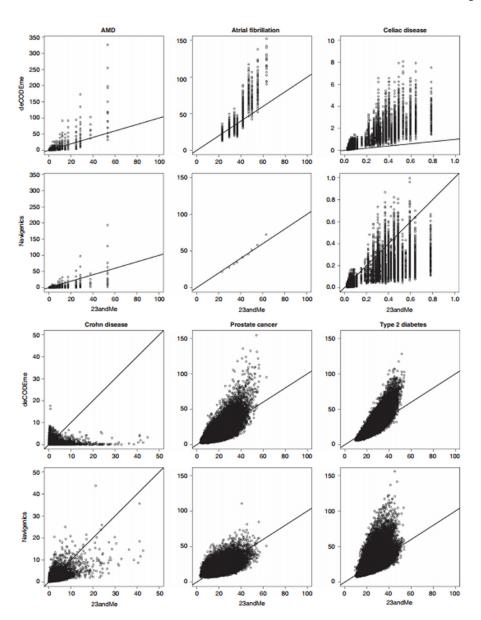
Figure 1. Predicted risks by 23andMe, deCODEme, and Navigenics for six multifactorial diseases

The figure shows the predicted risks for a hypothetical population of 100,000 individuals (see Materials and Methods section). The solid line indicates when predicted risks by deCODEme or Navigenics are the same as predicted risks by 23andMe. Mote than the ranges of the axes differ among the companies. AMD, age-related macular degeneration.

Table 1

Average population risks and number of SNPs used by 23andMe, deCODEme, and Navigenics in the prediction of risks for six multifactorial diseases

Diseases	Average population risk(%)			Number of SNPs		
	23andMe	deCODEme	Navigenics	23 and Me	deCODEme	Navigenics
Age-related macular degeneration	6.5	8	3.1	3	6	6
Atrial fibrillation	27.2	25	26	2	6	2
Celiac disease	0.12	1	0.06	4	8	10
Crohn disease	0.53	0.5	0.58	12	30	27
Prostate cancer	17.8	16	17	12	26	9
Type 2 diabetes	25.7	25	25	11	21	18

SNPs, single nucleotide polymorphisms.

Table 2

Area under the receiver operating characteristic curve for the prediction of six multifactorial diseases by 23andMe, deCODEme, and Navigenics

Diseases	23andMe	deCODEme	Navigenics
Age-related macular degeneration	0.76	0.81	0.82
Atrial fibrillation	0.58	0.62	0.58
Celiac disease	0.73	0.82	0.80
Crohn disease	0.76	0.80	0.77[a]
Prostate cancer	0.61	0.68	0.60
Type 2 diabetes	0.59	0.64	0.63[a]

[a] Calculated using an approximation of the formula described by Navigenics in their white paper (see Supplementary Materials and Methods online).

Table 3

Illustration of predictive ability using risk categories that 23andMe uses to classify disease risks

Diseases		23andMe			deCODEme			Navigenics		
		→	-	↑	→	-	↑	→	-	↑
Age-related macular degeneration	Patients	26.7	1.7	71.2	23.0	13.0	64.0	18.2	11.6	70.2
	NonPatients	66.7	2.1	31.2	69.2	13.2	17.6	66.5	10.5	23.0
Atrial fibrillation	Patients	53.1	14.4	32.5	34.1	33.1	32.9	52.6	14.2	33.2
	NonPatients	66.3	12.5	21.2	49.5	33.4	17.1	65.9	13.0	21.2
Celiac disease	Patients	42.1	0.0	57.9	22.9	2.5	74.6	24.6	4.9	70.5
	NonPatients	80.6	1.1	18.3	75.4	2.8	21.8	75.7	4.9	18.4
Crohn disease	Patients	31.2	10.1	58.7	39.7	10.9	49.4	26.7[a]	15.1[a]	58.3[a]
	NonPatients	69.7	10.9	19.4	83.8	5.5	10.7	65.6[a]	13.0[a]	21.5[a]
Prostate cancer	Patients	17.5	33.3	49.3	23.3	22.1	48.6	25.8	38.4	35.9
	NonPatients	29.7	36.3	34.0	54.1	21.7	24.2	38.4	38.8	22.9
Type 2 diabetes	Patients	23.0	43.5	33.5	29.3	26.9	43.8	31.3[a]	31.2[a]	37.5[a]
	NonPatients	33.5	43.7	22.8	48.4	27.2	24.5	48.4[a]	31.5[a]	20.1[a]

23andMe categorizes disease risks as decreased (↓), elevated (↑), and typical (-) risks if the risks of disease are lower than 20% below the average population risk, higher than 20% above the average population risk, and in between, respectively. Values are percentages.

[a] Calculated using an approximation of the formula described by Navigenics in their white paper (see Supplementary Materials and Methods online).

Table 4

Agreement among the three companies in assigning individual consumers to the same risk category, according to the risk categories used by 23andMe

Diseases	Assigned to the same risk category by all three companies			Assigned to the same risk category by two companies			Assigned to different risk categories
	↓↓↓	---	↑↑↑	↑↑- ↓↓-	--↑ --↓	↑↑↓ ↓↓↑	↑-↓
Age-related macular degeneration	52.3	0.5	15.2	6.1	6.0	12.5	7.4
Atrial fibrillation	42.4	6.7	16.7	27.3	5.7	1.2	0.0
Celiac disease	75.3	0.0	13.8	9.0	0.4	1.3	0.3
Crohn disease	51.8	0.2	3.5	13.8	3.7	19.9	7.2
Prostate cancer	15.6	4.5	13.5	29.4	21.7	6.5	9.0
Type 2 diabetes	22.2	7.8	14.7	24.1	23.1	3.2	5.0

23andMe categorizes disease risks as decreased (↓), elevated (↑), and typical (-) risks if the risks of disease are lower than 20% below the average population risk, higher than 20% above the average population risk, and in between, respectively. Values are percentages. For example, ↓↓↓ indicates the percentage of individuals that were decreased risk according to all three companies, and ↑-↓ indicates the percentage risks in the three different risk categories.