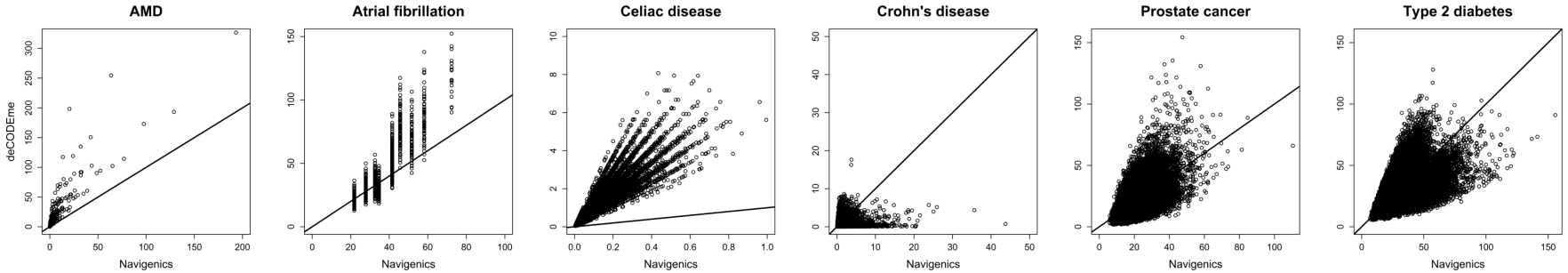
SUPPLEMENTAL INFORMATION

Variations in Predicted Risks in Personal Genome Testing for Common Complex Diseases

Rachel R.J. Kalf, MSc; Raluca Mihaescu, MD; Suman Kundu, MSc; Peter de Knijff, PhD; Robert C. Green, MD, MPH; A. Cecile J. W. Janssens, PhD

Figure S1. Predicted risks by deCODEme and Navigenics for six multifactorial diseases



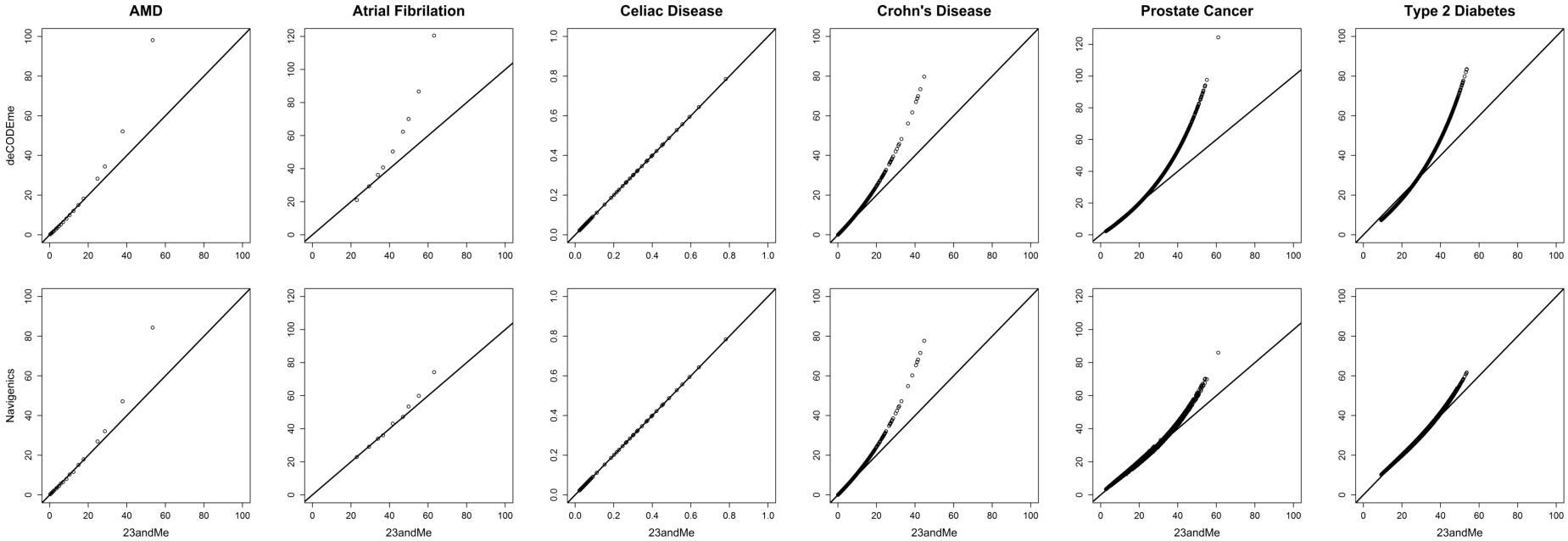
Legend: Predicted risks for a hypothetical population of 100 000 individuals (see Methods). The solid line indicates when predicted risks by deCODEme are the same as predicted risks by Navigenics. Note that the ranges of the axes differ between the companies.

SUPPLEMENTAL INFORMATION

Variations in Predictive Estimates in Personal Genome Testing for Common Complex Diseases

Rachel R.J. Kalf, MSc; Raluca Mihaescu, MD; Suman Kundu, MSc; Peter de Knijff, PhD; Robert C. Green, MD, MPH; A. Cecile J. W. Janssens, PhD

Figure S2. Predicted risks using the formulas by 23andMe, deCODEme and Navigenics when the same average population risks, odds ratios and allele frequencies were assumed



Legend: Predicted risks for a hypothetical population of 100,000 individuals were calculated using the formulas of the companies applied to the average population risks, odds ratios, allele frequencies and number of SNPs used by 23andMe. The solid line indicates when predicted risks by respectively deCODEme and Navigenics were the same as predicted risks by 23andMe.

SUPPLEMENTAL INFORMATION

Variations in Predicted Risks in Personal Genome Testing for Common Complex Diseases

Rachel R.J. Kalf, MSc; Raluca Mihaescu, MD; Suman Kundu, MSc; Peter de Knijff, PhD; Robert C. Green, MD, MPH; A. Cecile J.W. Janssens, PhD

Table S1. Overview of the single nucleotide polymorphisms and odds ratios used by 23andMe, deCODEme and Navigenics

SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	**deCODEme**	**Navigenics**
Age-related macular degeneration					
rs10490924	*10q26.13*	rs3750847	3.47*	3.47*	2.72/10.57
rs1061170	*1q31.3*	rs1061147	2.85*	Haplotype	3.10/6.30
rs1410996	*1q31.3*			Haplotype	3.16
rs2230199	*19p13.3*			1.69	1.70/2.60
rs547154	*6p21.33*	rs522162	1.88	1.88	3.13*
rs9332739	*6p21.33*			2.14	2.78
Atrial fibrillation					
rs10033464	*4q25*		1.39	Haplotype	1.39
rs13376333	*1q21.3*			1.13	
rs2200733	*4q25*		1.72	Haplotype	1.72
rs3807989	*7q31.2*			1.09	
rs3825214	*12q24.21*			1.14	
rs7193343	*16q22.3*			1.21	
Celiac disease					
rs1464510	*3q28*	rs9851967	1.22*		1.21
rs1738074	*6q25.3*			1.21	1.21
rs17810546	*3q25.33*			1.34	1.34
rs2187668	*6p21.32*		7.04	7.04	7.04
rs231779	*2q33.2*				1.24
rs2816316	*1q31.2*			1.41	1.41
rs3184504	*12q24.12*			1.19	1.19
rs6441961	*3p21.31*		1.32	1.21	
rs6822844	*4q27*	rs6840978	1.41	1.59	1.43*
rs917997	*2q12.1*			1.27	1.27
rs9811792	*3q25.33*				1.21
Crohn's disease					
rs1000113	*5q33.1*	rs7714584	1.33*		1.54/1.92
rs10045431	*5q33.3*			1.11	
rs10758669	*9p24.1*			1.12	
rs10761659	*10q21.2*		1.23/1.55		1.23/1.55
rs10883365	*10q24.2*	rs11190140	1.18*	1.18*	1.20/1.62
rs11175593	*12q12*				1.54

SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	**deCODEme**	**Navigenics**
rs11209026	*1p31.3*		2.92	Haplotype	
rs11805303	*1p31.3*	rs1004819* rs10889677**	1.56*	Haplotype**	1.39/1.86
rs11584383	*1q32.1*	rs12122721		1.18	1.18*
rs12521868	*5q31.1*			1.21	
rs1456893	*7p12.2*			1.20	1.20
rs1551398	*8q24.13*			1.08	1.08
rs17234657	*5p13.1*		1.16	1.16	1.54/2.32
rs1736135	*21q21.1*			1.18	1.18
rs17582416	*10p11.21*	rs4934724		1.16*	1.16
rs1793004	*11p15.1*			1.24	
rs2066843	*16q12.1*			1.37	
rs2066844	*16q12.1*		2.97		1.97/3.29
rs2066845	*16q12.1*		6.32		3.05/12.13
rs2066847	*16q12.1*		6.68		1.97/3.29
rs224136	*10q21.2*			1.67	
rs2241880	*2q37.1*	rs10210302	1.45/1.77	1.45	1.19/1.85*
rs2274910	*1q23.3*			1.14	
rs2301436	*6q27*			1.21	1.21
rs2476601	*1p13.2*	rs6679677		1.31*	1.31
rs2542151	*18p11.21*	rs1893217	1.15*	1.35	1.30/2.01
rs2872507	*17q12*			1.12	1.12
rs3764147	*13q14.11*			1.25	1.25
rs4263839	*9q32*			1.22	1.22
rs4958847	*5q33.1*			1.36	
rs6908425	*6p22.3*			1.21	
rs744166	*17q21.2*			1.18	1.18
rs762421	*21q22.3*			1.13	1.13
rs7746082	*6q21*			1.17	1.17
rs7927894	*11q13.5*				1.16
rs9286879	*1q24.3*			1.19	1.19
rs9858542	*3p21.31*	rs3197999	1.20*	1.17	1.09/1.84
Prostate cancer					
rs10086908	*8q24.21*			1.15	
rs10486567	*7p15.2*		1.19/1.37	1.12/1.19	1.23/1.33
rs10505483	*8q24.21*	rs16901979	1.44/2.17	1.79	1.79*
rs10896449	*11q13.3*			1.19/1.47	1.19/1.47
rs10934853	*3q2.3*			1.12	
rs10993994	*10q11.23*		1.23	1.24	
rs12621278	*2q31.1*	rs10207654	2.23	1.33*	

SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	**deCODEme**	**Navigenics**
rs1447295	*8q24.21*		1.43/2.23	1.53	1.43/2.23
rs1512268	*8p21.2*		1.19/1.39	1.18	
rs16902104	*8q24.21*			1.21	
rs17021918	*4q22.3*		1.20	1.11	
rs1859962	*17q24.3*		1.20	1.20	1.20
rs2660753	*3p12.1*			1.08	
rs2735839	*19q13.33*			1.12	
rs401681	*5p15.33*			1.07	
rs4430796	*17q12*		1.17	1.22	1.24/1.48
rs4962416	*10q26.13*				1.16 /1.49
rs5759167	*22q13.2*			1.16	
rs5945572	*Xp11.22*			1.23	
rs620861	*8q24.21*			1.14	
rs6465657	*7q21.3*			1.12	
rs6983267	*8q24.21*		1.25	1.26/1.58	1.26/1.58
rs7127900	*11p15.5*		1.24/1.42	1.22	
rs721048	*2p15*	rs2710646		1.15*	1.15
rs7679673	*4q24*			1.10	
rs8102476	*19q13.2*		1.12	1.12	
rs9364554	*6q25.3*			1.14	
Type 2 diabetes					
rs10010131	*4p16.1*	rs10012946	1.15/1.23*	1.11	1.03/1.19
rs10244051	*7p21.2*			1.06	
rs10830963	*11q14.3*			1.09	
rs10923931	*1p12*	rs2793831		1.13*	1.13
rs1111875	*10q23.33*		1.13	1.17	1.11/1.21
rs13266634	*8q24.11*		1.12	1.14	1.18
rs1387153	*11q14.3*		1.09		
rs1801282	*3p25.2*		1.14	1.14	1.30/1.53
rs2237892	*11p15.4*	rs2283228	1.29	1.29	1.24*
rs2383208	*9p21.3*	rs10811661	1.19	1.20	1.20/1.44*
rs2877716	*3q21.1*			1.12	
rs340874	*1q32.3*			1.07	
rs4402960	*3q27.2*		1.14	1.14	1.14
rs4430796	*17q12*			1.10	1.08/1.19
rs4607103	*3p14.1*				1.09
rs4607517	*7p13*			1.07	
rs5215	*11p15.1*	rs5219	1.14*	1.15	1.11
rs7578597	*2p21*			1.15	

SNP	Locus	Proxy SNP	Odds ratio		
			23andMe	**deCODEme**	**Navigenics**
rs7756992	*6p22.3*	rs4712523	1.12*	1.20	1.15/1.50
rs780094	*2p23.3*			1.06	
rs7903146	*10q25.2*	rs4506565	1.37	1.37	1.36/1.88*
rs7961581	*12q21.1*			1.09	1.09
rs8050136	*16q12.2*				1.17
rs864745	*7p15.1*			1.10	1.10
rs9300039	*11p12*				1.80/2.61
rs9494266	*6q23.3*				2.31

Legend: Values are allelic or genotypic odds ratios (heterozygous/homozygous) for the single nucleotide polymorphisms (SNPs) that are used by the companies. SNPs that were in linkage disequilibrium ($r^2 \geq 0.6$; SNP annotation and proxy search of the Broad Institute) are mentioned on the same line as reference SNP and proxy SNP. We selected the reference SNP as the SNP that was reported in the cited scientific studies; used by two of the three companies; or the first or most published SNP in the GWAS catalog. Haplotype indicates that the SNP was used to construct a haplotype (odds ratios not reported). Locus names were obtained from Ensembl.

*/** Odds ratio of the proxy SNP.

SUPPLEMENTAL INFORMATION

Variations in Predicted Risks in Personal Genome Testing for Common Complex Diseases

Rachel R.J. Kalf, MSc; Raluca Mihaescu, MD; Suman Kundu, MSc; Peter de Knijff, PhD; Robert C. Green, MD, MPH; A. Cecile J.W. Janssens, PhD

SUPPLEMENTARY MATERIALS AND METHODS

The Supplementary methods describe in more detail how the datasets were constructed and which sources of input data (odds ratios and genotype frequencies) were used for each research question. We also report our efforts to verify the risk calculations.

Simulated data

Construction of genotype data

Simulated datasets were constructed using a modeling procedure that has been validated and described in more detail elsewhere.[1,2] 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 the genotype or allele frequencies match pre-specified input values. The input values for the frequencies were obtained from Hapmap, from the scientific publications cited on the websites of the companies or directly from those websites, depending on the research question. Which source was used for each research question is specified below. When input values were allele frequencies, Hardy-Weinberg equilibrium was assumed to obtain genotype frequencies.

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.[3-5] 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 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).[6-8] For the calculation of the likelihood ratios and relative risks that are needed to compute predicted risks, genotype and allele frequencies were obtained from Hapmap release #24 for 23andMe, from the cited scientific studies for deCODEme and from the company's website for Navigenics. For three SNPs used by 23andMe genotype frequencies were not available in Hapmap and hence were calculated from the odds ratios and likelihood ratios in available reports (see below). All risks were calculated for Caucasian men.

Although the methods of the three companies require the same input parameters, the formulas used for the exact calculation of risks had notable differences. 23andMe and deCODEme transformed genotype odds ratios of single SNPs into likelihood ratios, representing the odds of disease for each genotype relative to the average odds. To compute predicted risks, 23andMe then multiplied the likelihood ratios of single genotypes by the average odds of disease and converted the odds into risks, whereas deCODEme multiplied the likelihood ratios by the average risk of disease. Navigenics did not transform the individual genotype odds ratios into likelihood ratios but into relative risks, and calculated the relative risks for all possible genotype combinations. To compute individual predicted risks, they divided the relative risk of each genotype combination by the average *relative* risk before multiplying these by the average *absolute* risk of disease. Given that for each SNP the reference genotype for the calculation of the relative risk was the genotype with the lowest risk of disease, this strategy is in essence similar to multiplying all relative risks by the theoretically lowest possible risk. Because Navigenics calculates the relative risks for all possible genotype combinations, the method requires substantial computer working memory. On a standard computer the method could run out of memory when the number of SNPs is higher than 14, which was the case for Crohn's disease and type 2 diabetes. For these diseases, we obtained an approximation of the lowest possible risk by dividing the population

disease risk by the average of the relative risks for each genotype combination in our population.

Verification of risk calculations

To verify whether we were applying the methods of the companies accurately, we first attempted to reproduce the risks presented in the sample reports that are available on their websites as well as to reproduce the risks predicted for two researchers who had their DNA tested by each of these companies (RG and PdK). For the six diseases under study, we were able to verify a total of 48 predicted risks; 18 for 23andMe and deCODEme and only 12 for Navigenics because we only had information from the sample report and the report of one researcher. We exactly reproduced 31 of the 48 predicted risks, and found absolute differences smaller than 1% for 12 risks and differences up to 5% for 5 risks. These differences were mainly explained by the fact that for a few SNPs we could not retrieve the exact same odds ratios used by the companies.

Apart from verifying whether we accurately could reproduce risks from the available reports, we also verified for individual SNPs whether we could reproduce the likelihood ratios (23andMe and deCODEme) and odds ratios (Navigenics) that were presented in the reports and used in the calculations. The 44 SNPs tested by 23andMe, 97 SNPs by deCODEme and 72 SNPs by Navigenics generated 132, 302 and 216 likelihood ratios and odds ratios for included SNPs and haplotypes (for deCODEme). For 80 (61%), 174 (58%) and 174 (81%), at least one in each SNP, we knew the exact likelihood ratios used by 23andMe and deCODEme and the exact odds ratios that Navigenics used to calculate relative risks from the available reports. More information was available for Navigenics, because the odds ratios of the homozygous genotypes were given for all SNPs in each report. Using odds ratios and genotype frequencies from the literature, we could reproduce 85% (68/80), 94% (163/174)

and 98% (170/174) of the likelihood ratios and odds ratios with an absolute difference of 0.01 or smaller. For 23andMe, we were able to reproduce only 85% with an absolute difference of 0.01 or smaller, because for several SNPs the genotype frequencies were not in Hapmap release #24 and because, more often than for the other companies, the odds ratios found in the literature did not exactly reproduce the likelihood ratios reported in the available reports. Yet, absolute differences were 0.05 or larger for only three likelihood ratios of 23andMe, one of deCODEme and one odds ratio of Navigenics. Because we knew at least one likelihood ratio or odds ratio for each SNP, we assumed we could also correctly obtain the remaining likelihood ratios and odds ratios. For 23andMe and deCODEme, we could reproduce an additional 9% (7/80) and 3% (5/174) of the likelihood ratios by using likelihood ratios from the available reports and odds ratios from the literature. These efforts suggest that we managed to reconstruct the prediction methods of the companies. The minor differences in likelihood ratios and odds ratios might have affected the exact calculation of the predicted risks, but they were unlikely large enough to influence the main findings of the study.

Data analysis

Comparison of predicted risks

To compare predicted risks among the 3 companies, we constructed one large dataset with genotypes for all 113 SNPs tested by the three companies for all six diseases. These 113 SNPs remained from the total of 213 SNPs (44 for 23andMe, 97 for deCODEme and 72 for Navigenics) after excluding duplicates (n=73) and SNPs in linkage disequilibrium (LD; $r^2 \geq 0.6$; n=27). If SNPs were in LD, we selected the SNP that was 1) reported in the cited scientific studies; 2) used by two of the three companies; or 3) the first or most published in the GWAS catalog.[9] Genotype frequencies were obtained from Hapmap release #28, except

for two SNPs that were not available in Hapmap we used the frequencies reported on the website of Navigenics.

Comparison of predictive ability

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 separately. The predictive ability was quantified by the area under the receiver operating characteristic curve (AUC).[10] AUC values range from 0.5 (random prediction) to 1.0 (perfect prediction). 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 AUC, disease status was randomly assigned to individuals based on 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.[1]

Note that to assess the predictive ability we constructed genotype datasets for each company separately based on the genotype frequencies they had used for the calculation of their predicted risks, instead of using a single set of frequencies for all companies, like we did for the first research question. Constructing separate datasets for each company based on their own input values ensures that all risk models are perfectly calibrated. Had we used e.g. Hapmap #24 to construct one large genotype dataset, then the risk models of 23andMe were expected to be better calibrated than those of deCODEme and Navigenics as 23andMe used Hapmap #24 genotype frequencies in the calculation of the disease risks. In that case, differences in AUC would have been in part due to differences in calibration. While this approach yields a valid comparison of the genetic tests, the exact AUC values should be

interpreted with caution as they are likely overestimated. For all three companies, external validation of the risk models in an independent unselected population will likely show lower AUC values and suggest poorer predictive ability than presented in this study.

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).[6]

Comparison of risk categories

Finally, we assessed the agreement between the companies in classifying each individual to the same risk category. We used the original large dataset, constructed for the comparison of predicted risks between 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.[11]

SUPPLEMENTAL REFERENCES:

1. 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(7):395-400.

2. 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(12):1270-1274.

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

4. deCODEme. SNP selection and risk calculation. Available at: http://demo.decodeme.com/health-watch-information/risk-calculation. Accessed July 1, 2012.

5. Navigenics. The science behind the Navigenics service. Available at: http://www.navigenics.com/static/pdf/Navigenics-TheScience.pdf. Accessed July 1, 2012.

6. 23andMe. Available at: https://www.23andme.com/. Accessed January 1, 2012.

7. deCODEme. Available at: http://www.decodeme.com/. Accessed January 1, 2012.

8. Navigenics. Available at: http://www.navigenics.com/. Accessed January 1, 2012.

9. Hindorff LA, MacArthur J, Morales J, et al. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed July 1, 2012.

10. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.

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