August 2007 · Vol. 9 · No. 8

The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases

A. Cecile J. W. Janssens, PhD[1], Ramal Moonesinghe, PhD[2], Quahne Yang, PhD[3], Ewout W. Steyerberg, PhD[1], Cornelia M. van Duijn, PhD[4], and Muin J. Khoury, PhD[2]

Purpose: Single genetic variants in multifactorial disorders typically have small effects, so major increases in disease risk are expected only from the simultaneous exposure to multiple risk genotypes. We investigated the impact of genotype frequencies on the clinical discriminative accuracy for the simultaneous testing of 40 independent susceptibility genetic variants. **Methods:** In separate simulation scenarios, we varied the genotype frequency from 1% to 50% and the odds ratio for each genetic variant from 1.1 to 2.0. Population size was 1 million and the population disease risk was 10%. Discriminative accuracy was quantified as the area under the receiver-operating characteristic curve. Using an example of genomic profiling for type 2 diabetes, we evaluated the area under the receiver-operating characteristic curve when the odds ratios and genotype frequencies varied between five postulated genetic variants. **Results:** When the genotype frequency was 1%, none of the subjects carried more than six of 40 risk genotypes, and when risk genotypes were frequent (≥30%), all carried at least six. The area under the receiver-operating characteristic curve did not increase above 0.70 when the odds ratios were modest (1.1 or 1.25), but higher genotype frequency increased the area under the receiver-operating characteristic curve from 0.57 to 0.82 and from 0.63 to 0.93 when odds ratios were 1.5 or 2.0. The example of type 2 diabetes showed that the area under the receiver-operating characteristic curve did not change when differences in the odds ratios were ignored. **Conclusions:** Given that the effects of susceptibility genes in complex diseases are small, the feasibility of future genomic profiling for predicting common diseases will depend substantially on the frequencies of the risk genotypes. ***Genet Med* 2007:9(8):528–535.**

Key Words: *genetic, genomics, predictive test, clinical validity*

Advances in genetics and genomics bear the promise of individualized medicine, in which health interventions, lifestyle recommendations, and medication prescriptions are targeted based on individuals' genetic susceptibilities.[1,2] Genome-based individualized health interventions for common chronic diseases, such as type 2 diabetes, asthma, osteoporosis, and cardiovascular disease, will be based on the simultaneous testing of multiple genetic variants, so-called genomic profiling, because single genetic variants each confer only a minor increase in the risk of disease. To date, it is unclear whether genomic profiling can be useful for the prediction of chronic diseases.

The usefulness of genomic profiling can be examined in large-scale cohorts of individuals who have been extensively genotyped and for whom prospective follow-up data on phenotypes and disease outcomes are available,[3,4] or it can be estimated in case-control studies if conducted in population-based settings where the overall disease risk is known. Nevertheless, major gaps currently exist in our knowledge base in this area, which may have to wait years if not decades of human genome epidemiologic investigations. Awaiting these data, we investigated the clinical validity of future genomic profiling using hypothetical but plausible simulated data.

Using simulation studies, we investigated the discriminative accuracy of genomic profiling, which is the extent to which genomic profiles can distinguish between those who will or will not develop disease. We demonstrated that the simultaneous testing of 20 weak susceptibility genes may yield a level of discriminative accuracy comparable with that of total serum cholesterol testing for the prediction of coronary heart disease and neuropsychological testing for the prediction of Alzheimer disease in asymptomatic individuals.[5–7] The simultaneous test-

From the [1]Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands; [2]National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia; [3]National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia; [4]Department of Epidemiology and Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands.

Disclosure: The authors declare no conflict of interest.

A. Cecile J.W. Janssens, PhD, Erasmus MC, University Medical Center Rotterdam, Department of Public Health, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands;; e-mail: a.janssens@erasmusmc.nl.

Submitted for publication March 27, 2007.

Accepted for publication May 15, 2007.

DOI: 10.1097/GIM.0b013e31812eece0

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

ing of hundreds of susceptibility genes yielded an even better discriminative accuracy, which could be useful for the identification of individuals at increased genetic risk of disease in asymptomatic populations, but unlikely for the presymptomatic diagnosis of disease such as possible with genetic testing for Huntington disease.[8] We also demonstrated that a major proportion of common chronic diseases can be explained by only a few common genotypes or by a great many rare genotypes,[9] which suggests that the identification of such common genotypes is important in the feasibility of genomic profiling.

As single genetic variants in multifactorial disorders typically have small effects (as measured in terms of odds ratios [ORs] or risk ratios, major increases in disease risk are expected only from the simultaneous exposure to multiple risk genotypes. The size of the group that has an increased risk of disease depends on the frequencies of the individual risk genotypes: when risk variants are infrequent, substantial increases in risk may only be expected for very small subgroups of the population tested.[10] The aim of the present analysis was to investigate in a simulation study the impact of genotype frequencies on the clinical validity of genomic profiling for the prediction of chronic diseases.

MATERIALS AND METHODS

Modeling strategy

A genomic profile is defined by the genotype status on multiple genetic markers. The clinical validity of genomic profiling is determined from the distributions of disease risks in subjects who will develop the disease and those who will not. The disease risk associated with each profile, i.e., the predictive value of the profile, is determined by the presence or absence of risk genotypes on the genetic variants in the profile. For additive, multiplicative, or other forms of synergistic joint effects at multiple loci on disease risk, the more risk genotypes there are, the higher is the disease risk associated with the genomic profile. To obtain the distributions of disease risks for the calculation of the clinical validity, we need to specify: (1) the genomic profiles of all individuals, (2) the disease risks associated with the profiles, and (3) the disease status of all individuals. These are modeled in three subsequent steps, for each combination of the study parameters (genotype frequencies and odds ratios) separately.

Modeling genomic profiles

In the modeling study, we assumed that at each individual gene, there are two genotypes, one of which was associated with an increased risk of disease—referred to as the risk genotype— and the other with the referent or baseline risk. Genetic variants were modeled to be independent, i.e., no linkage disequilibrium existed between them. For the construction of the genomic profiles, we first created a vector for each genetic variant with as many copies of the two genotypes as indicated by their frequencies and the sample size. Assuming that the alleles at individual loci segregate independently, we then sampled, randomly without replacement, for each subject a genotype from each vector.

Modeling disease risks associated with genomic profiles

Next, we calculated the disease risks associated with the genomic profiles using Bayes' theorem. Bayes' theorem states that the posterior odds of disease are obtained by multiplying the prior odds by the likelihood ratio (LR) of the test result, or here with the LR of the genomic profile.[11] The previous odds are calculated from the population risk of disease:

$$\text{Prior odds} = \frac{\text{Population risk}}{1 - \text{Population risk}}$$

and the posterior odds are converted back into probabilities:

$$\text{Probability} = \frac{\text{Odds}}{1 + \text{Odds}},$$

which is the disease risk associated with the genomic profile.

The LR of a genomic profile can be obtained by multiplying the LRs of the single genotypes because we assumed that genotypes are transmitted independently from one generation to the next and that the joint risks of multiple genotypes follow a multiplicative risk model without any interactions.[12,13] The LR of a single genotype is the percentage of the genotype among subjects who will develop the disease divided by the percentage of the genotype among subjects who will not develop the disease. The LR can be calculated from a 2×2 table (disease status $\times$ genotype status), when the disease risk, genotype frequency, and the risk ratio are known. As risk ratio, we used the OR in our analyses because many genetic association studies are (nested) case-control studies, which yield ORs as risk estimates. Moreover, calculation of disease risks based on relative risks can yield disease risks for the profiles that can be greater than 100%. This problem, which easily can occur when combining 40 genetic factors with larger effects (e.g., relative risk of 2), does not occur when disease risks are calculated from ORs. This procedure has been described previously in more detail.[8]

Modeling disease status

To model disease status, we used a procedure that compares the disease risk of each subject to a randomly drawn value between 0 and 1 from a uniform distribution.[14] A subject was assigned to the group who will develop the disease when the disease risk was higher than the random value and to the group who will not develop the disease when the risk was lower than the random value. This procedure ensures that for each genomic profile, the percentage of people who will develop the disease equals the disease risk associated with that profile, when the subgroup of individuals with that profile is sufficiently large.

Statistical analysis

Although it is possible to obtain an analytic solution for the simple scenarios, we used simulation for greater flexibility. We simulated genomic profiles and disease status for 1 million subjects. In all scenarios, the population risk of disease was

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

Table 1
Genetic variants for type 2 diabetes susceptibility: data obtained from meta-analyses or pooled analyses of large-scale studies

	Published data			Data used in this study		
Genetic polymorphism	At-risk variant	Variant frequency	OR (CI)	At-risk genotype(s)	Variant frequency	OR
PPARγ P12A[17]	PP	78%	1.40 (1.12–1.76)	PP	78%	1.40
KCNJ11 E23K[18]	KK	13%	1.44 (1.17–1.78)	KK	13%	1.36[a]
	EK	46%	1.12 (1.01–1.23)			
CAPN10 SNP44 T(−g4841)C[19]	C allele	16%	1.17 (1.02–1.34)	TC/CC	29%[a]	1.19[a]
HNF1A A98V[20]	V allele	3%	1.31 (1.08–1.59)	AV/VV	6%[a]	1.31[a]
TCF7L2[21]	TT	7%	2.11 ($P = 5.16^{a}10^{-10}$)	TT	47%	1.51[a]
	GT	40%	1.42 ($P < 10^{-16}$)			

OR, odds ratio; CI, 95% confidence intervals.
[a]To obtain the data, we constructed a disease status × genotype status table based on the published genotype frequencies, odds ratios, and a disease risk of 33%.[16] We assumed that allele and genotype frequencies were in Hardy-Weinberg equilibrium. When odds ratios were published for the risk allele, we assumed a multiplicative effect associated with the presence of two copies of the risk allele. Heterozygous and homozygous risk genotypes were then combined and a composite OR was calculated.

10% and the genomic profiles consisted of 40 genetic variants. We recognize that the choice of these numbers is rather arbitrary, but it does help to illustrate the trend of combining genetic variants to predict a common disease in the population. To single out the effect of genotype frequencies, we first assumed that all genetic markers had the same genotype frequencies and the same effect sizes. This implies that genomic profiles with the same number of risk genotypes had the same associated disease risks. The frequencies of the risk genotypes varied from 1% to 50% and the ORs from 1.1 to 2.0. This range of ORs is based on statistically significant pooled ORs from 50 meta-analyses on various gene-disease associations.[15]

Combining 40 genetic variants with two genotypes yields a test with 2^{40} different profiles. We proposed that genomic profiling can be considered as a continuous test when the profiles are expressed by their associated disease risks.[10] Clinical validity was investigated as the discriminative accuracy, i.e., the extent to which genomic profiling can discriminate between subjects who will develop the disease and those who will not, which is for continuous variables indicated by the area under the receiver-operating characteristic curve (AUC). The AUC indicates whether a test is useful to identify individuals who are at increased risk of disease (screening; e.g., AUC 0.75–0.80) or to diagnose a disease before the onset of symptoms (presymptomatic diagnosis; AUC >0.99). The AUC shows the sensitivity (true positive rate) and specificity (1 − true negative rate) of the test when the profiles are categorized in two groups based on the disease risks associated with the profiles. The optimal combination of sensitivity and specificity is determined by external factors, such as the medical, psychological, and financial costs of false-positive and false-negative test results. The AUC is calculated from the distribution of disease risks of subjects who will develop the disease and of those who will not. The AUC was obtained as the *c* statistic by the function *somers2* in the Hmisc library of R software.[14]

Next, we modeled a more realistic scenario in which effect sizes and genotype frequencies differed between the genetic markers. As a result, genomic profiles with the same number of risk genotypes varied in the associated disease risks depending on which genotypes were present in the profile. We investigated the clinical validity of genomic profiling for the prediction of type 2 diabetes based on the simultaneous testing of the five susceptibility genes (Table 1). Population lifetime disease risk was considered as 33%.[16] AUCs were calculated for the discriminative accuracy of disease risks, which takes into account differences in the ORs of the genotypes, and of the number of risk genotypes, which does not take into account these differences.

All simulations were replicated 50 times to obtain more robust estimates for the size of the population at risk, for the disease risks associated with the genomic profiles, and for the estimates of the AUC. Average estimates of the 50 replications are presented. Because we simulated populations of 1 million subjects, confidence intervals around these estimates were so small (e.g., 95% confidence interval of AUC estimates was $<\pm 0.01$) that these could not be made visible in the graphs. All analyses were performed using R software version 2.5.0 (www.R-project.org; accessed May 2007).

RESULTS

Distributions of genomic profiles and disease risks

Figure 1 shows the distributions of the genomic profiles represented as the number of risk genotypes in the profiles. When the frequencies of all risk genotypes were 1%, two thirds of the population had no risk genotypes and one fourth had only one risk genotype of 40 genetic variants. None of the subjects had more than six risk genotypes, which means that the probability of having seven or more risk genotypes was lower than 1 per million. Yet when the frequency of the risk genotypes was 30%, all individuals had at least one risk genotype, and when the frequency was 50%, all had at least five. The distribution of the number of risk genotypes in the total population was indepen-

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

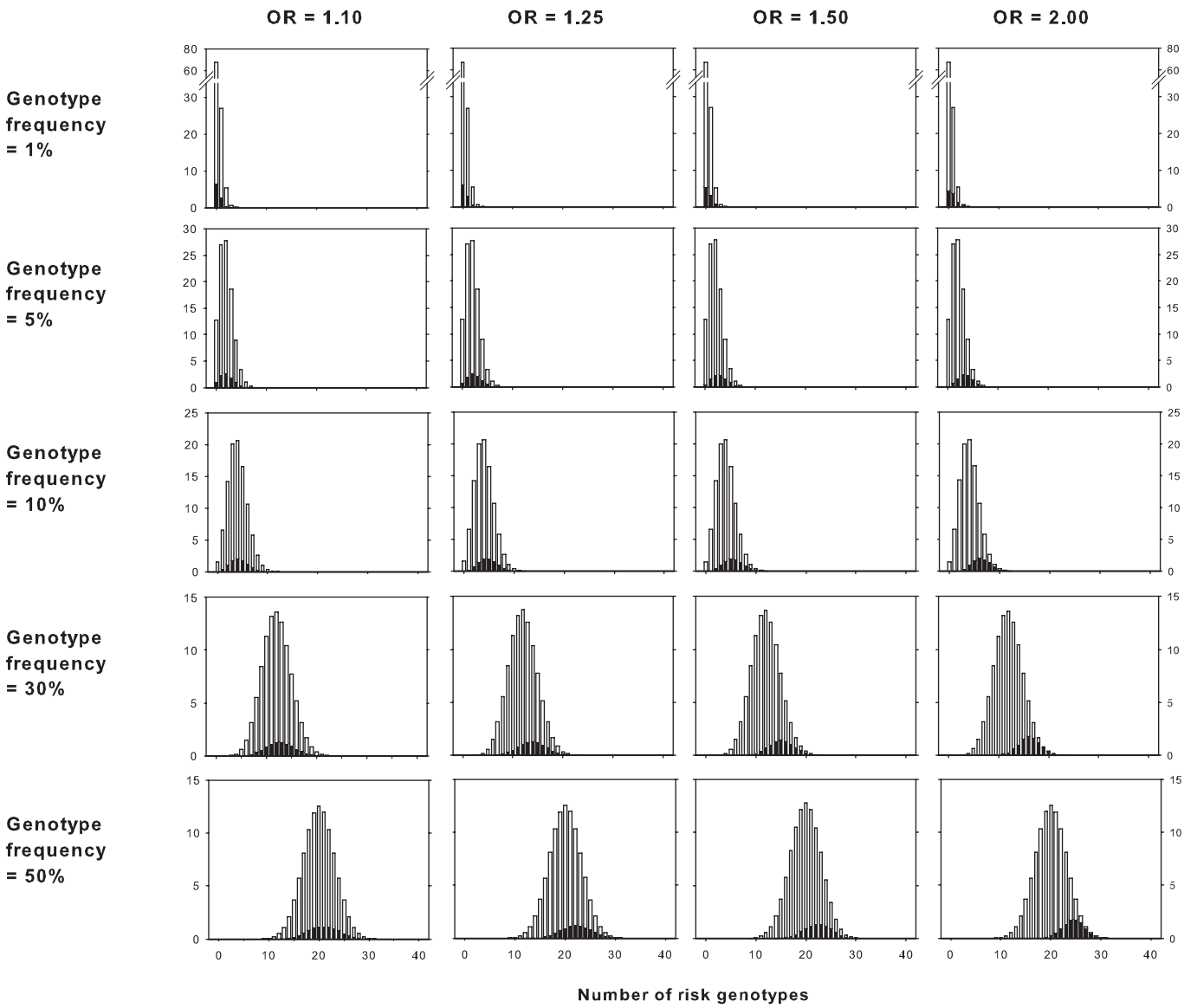

Fig. 1. Distributions of the genomic profiles in the total population as a function of the frequency and effects of the single risk genotypes. Genomic profiles were defined by 40 susceptibility genes, the population disease risk was 10%, and the population size was 1 million. The figures are stacked bar charts. The black parts of the bars indicate the percentage of the total population that will develop the disease and the white parts the percentage that will not develop the disease. The *x*-axis indicates the number of risk genotypes out of 40 genes in the genomic profile. The *y*-axis indicates the percentage of persons in the total population, e.g., 67% of the total population carries no risk genotypes when the genotype frequency is 1%.

dent of the disease risk in the total population (data not shown) and of the ORs of the risk genotypes (Fig. 1).

The dark bars in Figure 1 represent the distribution of the number of risk genotypes for individuals who will develop the disease and the white bars the distribution for those who will not. When the ORs were low (OR = 1.1 or 1.25), subjects who will develop the disease had approximately the same distribution of the number of risk genotypes compared with subjects who will not develop the disease, in terms of the shape and location of the distributions. This is explained by the modest effect sizes of the individual variants. When the ORs were 1.5 or higher, those who would develop the disease tended to have genomic profiles with more risk genotypes than those who would not develop the disease.

As expected, higher disease risks were obtained when ORs of the single genetic variants were higher (Fig. 2). When the OR was 1.1, genomic profiling at best resulted in a threefold increase in the disease risk when the genotype frequency was 50%, but this risk increase only applied to <1 in 100,000 individuals even when the risk genotypes were common. When the ORs of each genetic variant were 1.5 or 2.0, a larger proportion of the population had at least a fivefold increase in disease risk when the risk genotypes were common (5.0% when OR was 1.5 and 5.4% when OR was 2.0; Fig. 2).

Discriminative accuracy

Genomic profiling had low discriminative accuracy (AUC <0.70) when the ORs of the 40 genetic variants were low (1.1

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

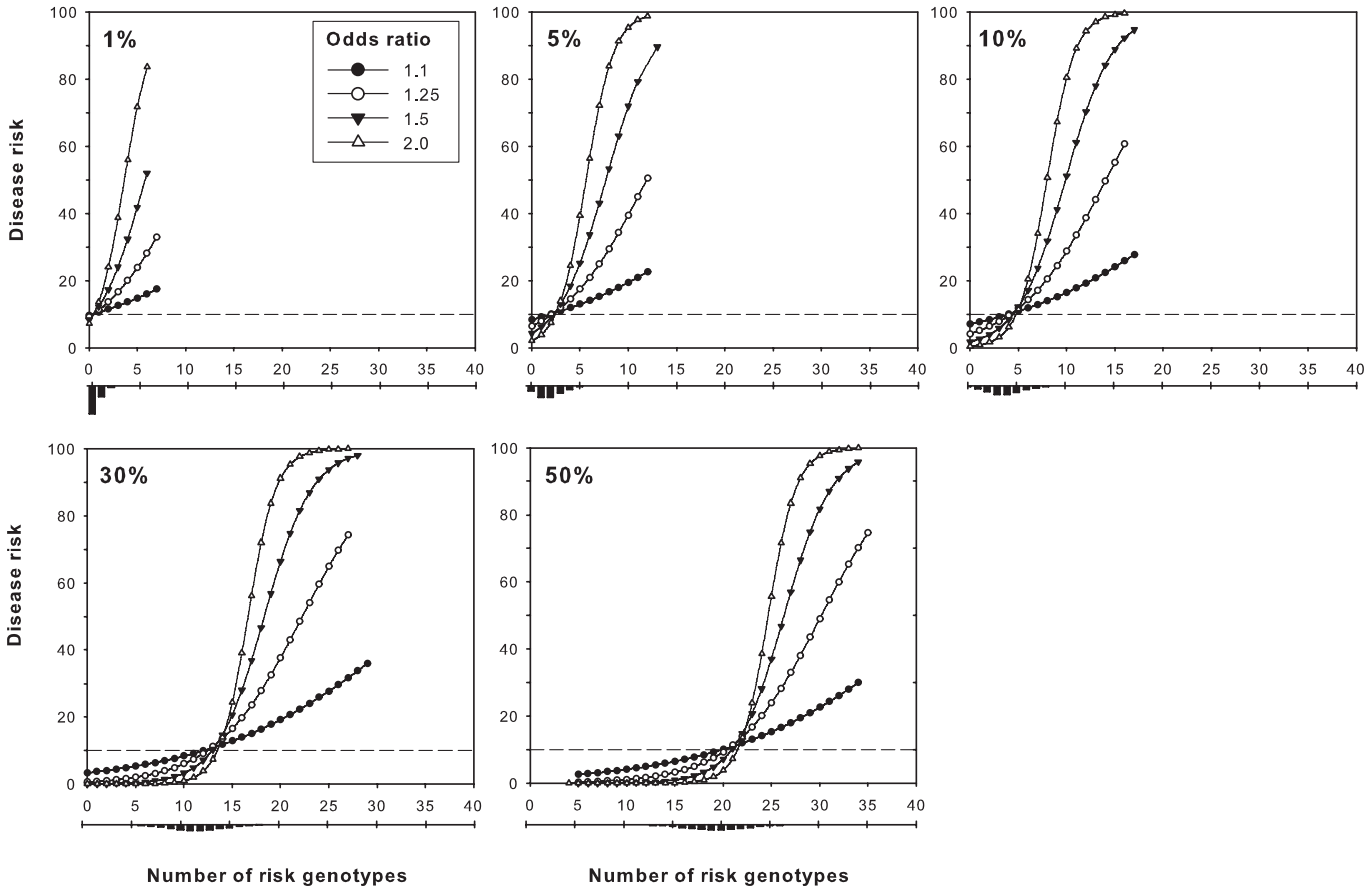

Fig. 2. Disease risks associated with genomic profiles resulting from the simultaneous testing of 40 susceptibility genes. Genomic profiles were defined by the number of risk genotypes, based on the simultaneous testing of 40 genes that all have the same genotype frequencies and odds ratios, as indicated. Population disease risk was 10% (indicated by the dashed line). The *x*-axis indicates the number of risk genotypes of 40 genes tested. The lower *x*-axis presents the distribution of the number of risk genotypes in the total population, which corresponds with the distributions in Figure 1. Population size was 1 million. The lines in the graphs end when the frequency of the number of risk genotypes was less than 1 in 1 million.

or 1.25; Fig. 3). Discriminative accuracy was also low when the profiles include genetic variants that were stronger predictors of disease but which had rare risk genotypes: when the frequency of the risk genotype was 1%, the AUC was 0.57, when ORs of the individual variants were 1.5 and 0.63 when their ORs were 2.0. The AUC increased substantially to 0.80 (OR = 1.5) and 0.92 (OR = 2.0) when the frequencies of the risk genotypes were 30%, but there was only minimal further improvement in the AUC when the frequencies increased to 50% (0.82 for OR = 1.5 and 0.93 for OR = 2.0).

Example

Figure 4 shows the disease risks and AUC for genomic profiling of type 2 diabetes in which the genetic variants varied in the effect sizes and frequencies of the risk genotypes. Differences in effect sizes resulted in variation in the type 2 diabetes risk within subgroups of individuals who had genomic profiles with the same number of risk genotypes, except for those whose profiles included none (6.7% of the population) or all five risk genotypes (0.1%). Forty percent of the population had genomic profiles that included two risk genotypes. These individuals had disease risks that were around the population risk of 33%. The AUC of genomic profiling based on the disease risk did not differ from that based on counting of the number of risk genotypes (both AUC = 0.58).

DISCUSSION

These results show that the discriminative accuracy of genomic profiling using 40 low-risk genetic variants was low when genotype frequencies were low, even when the ORs were 2.0. The discriminative accuracy was substantially higher when risk genotypes with modest effects were more frequent. Furthermore, the example of type 2 diabetes, in which ORs and genotype frequencies varied between the five postulated susceptibility genetic variants (ORs ranging from 1.19. to 1.51) showed that the AUC was rather low and did not change when differences in the ORs were ignored.

Three methodological issues need to be addressed. First, we found that none of the subjects had more than six risk genotypes of 40 genetic variants when the genotype frequency was 1%, and all subjects had at least five risk genotypes when the genotype frequency was 50%. This does not mean that the other combinations (e.g., 39 risk genotypes of 40) do not occur, but rather that they are extremely unlikely: they were not represented by chance in a population of 1 million subjects.

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

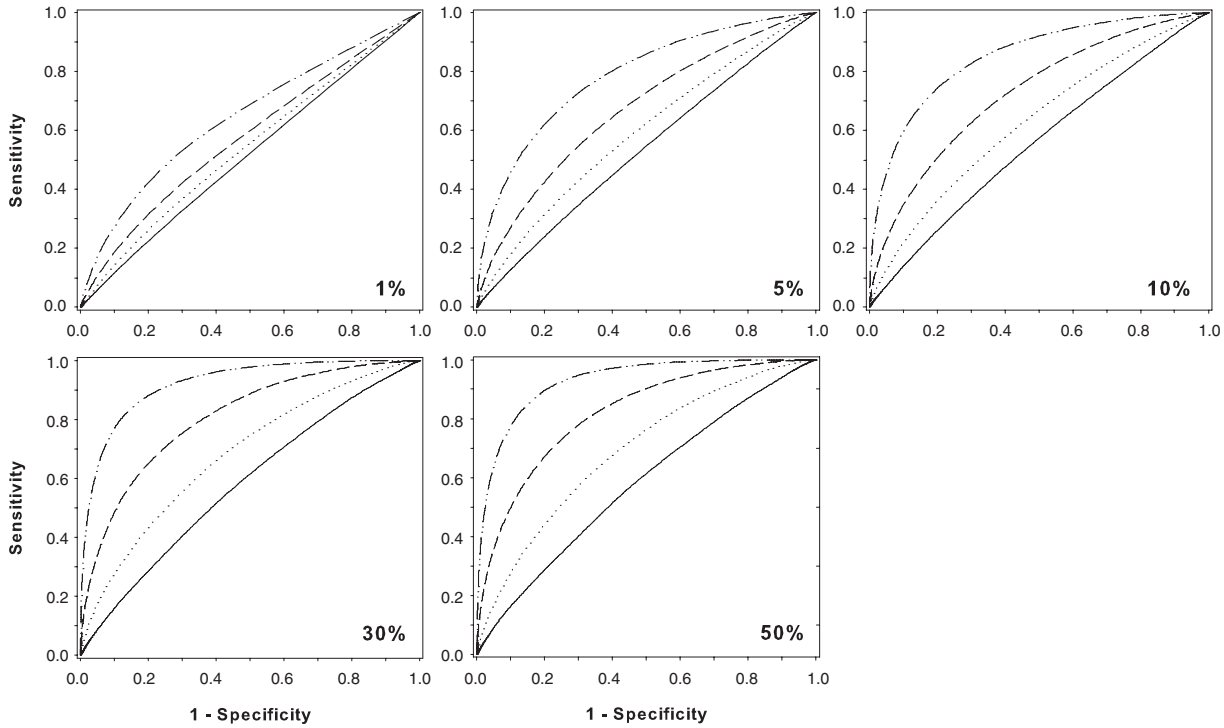

Fig. 3. Discriminative accuracy of genomic profiling as a function of the frequency and effects of the single risk genotypes. Genomic profiling was based on the testing of 40 genes that all have the same genotype frequencies and odds ratios, as indicated. Areas under the receiver-operating characteristic curve (AUCs) accompany the distributions of Figure 1. The values of the AUCs refer to the lines from left to right representing odds ratios of 2.0 (dashed-dotted line), 1.5 (dashed line), 1.25 (dotted line), and 1.1 (solid line). Population disease risk was 10%.

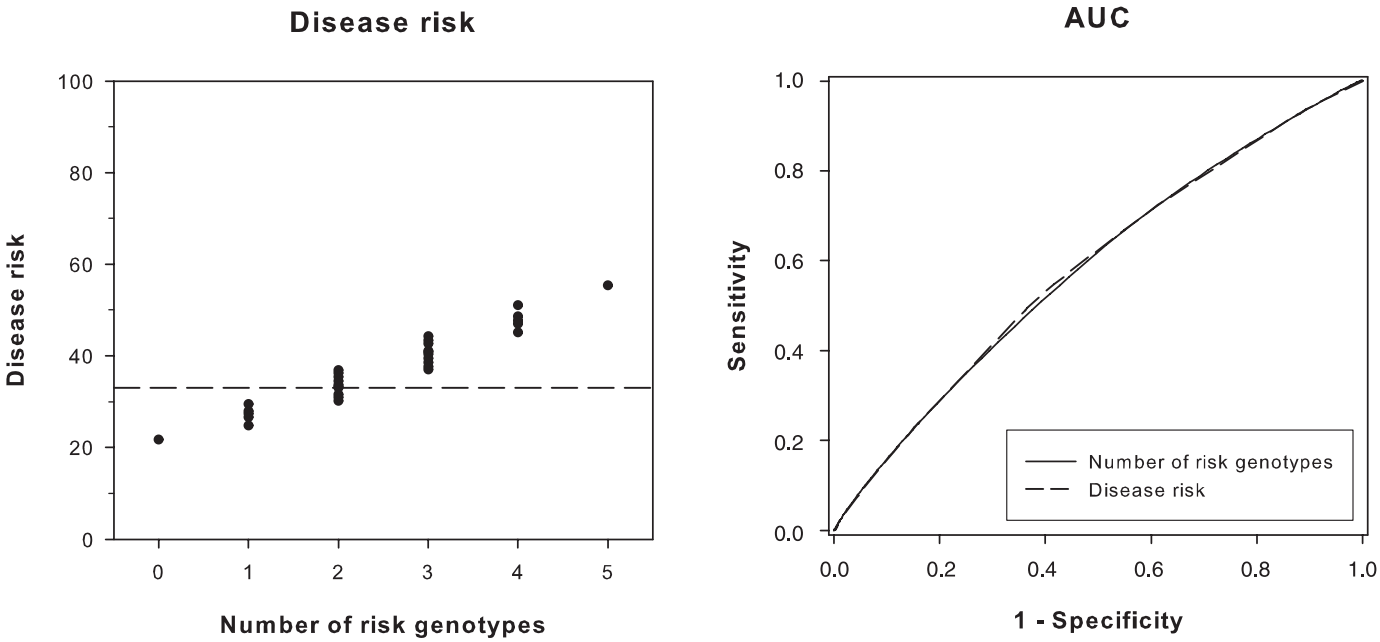

Fig. 4. Disease risks and discriminative accuracy of genomic profiling for predicting type 2 diabetes using five susceptibility genes. Population disease risk was 33% (dashed line).[16] AUC, area under the receiver-operating characteristic curve.

Second, we assumed that common diseases are caused by many common variants, each conveying only minor increases in disease risk, and investigated the role of genotype frequencies within a small range of effect sizes. Yet, under the common disease–rare variant hypothesis, multiple rare variants, each being a sufficient causal factor, may be combined into a larger category of variants with strong effects. This scenario, provided that the rare variants combined are not too rare, may be an

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

exception to our results. Third, similar to others, we assumed a multiplicative model to calculate the probability of disease for the genomic profiles.[12,13] When the ORs are the same for all variants, then the discriminative accuracy is determined by the number of risk genotypes in the genomic profile rather than by their ORs. This means that the discriminative accuracy will be the same under the assumptions of additive and multiplicative risk models. Yet, the disease risks associated with the genomic profiles do depend on the ORs of the assumptions of the risk model and will likely be lower and closer to the population disease risk when an additive risk model is assumed. Fourth, we did not consider gene-gene interaction effects in this report, as there is an infinite number of ways in which 40 genetic variants can interact. Depending on their strength and direction, interactions may or may not improve the discriminative accuracy of genomic profiling and change the disease risks associated with the profiles. Our model can be easily extended to include joint effects, but large-scale studies and meta-analyses should still demonstrate the role of joint effects to warrant evaluation of their contribution to clinical usefulness of genomic profiling.

When all risk genotypes were stronger risk factors of disease (OR = 1.5 or 2.0), which is not a very realistic scenario in complex diseases, discriminative accuracy was still poor (AUC <0.70) when the risk genotypes were rare, but excellent (AUC >0.80) when they were common. Frequent risk genotypes with ORs of 1.5 even yielded a test with higher discriminative accuracy than the same number of risk genotypes with ORs of 2.0 but that were rare. Higher frequency of the risk genotype also yielded a higher AUC when the odds ratios were low (OR = 1.1 or 1.25), but the discriminative accuracy remained poor. We previously demonstrated that genomic profiling by up to 400 susceptibility genes with ORs of 1.1 had a poor discriminative accuracy, but that profiling by 80 common variants with ORs of 1.25 could be useful for the identification of at-risk groups (AUC >0.80).[8]

In addition, the example of type 2 diabetes showed that there was no difference in the discriminative accuracy when genomic profiling was considered by counting the number of risk genotypes in each profile or by calculating the associated disease risks. Simply counting the number of risk genotypes in the profiles ignores that different genetic variants have different effect. The finding that the two approaches had the same AUC suggests that the differences in the effect sizes of the genetic variants in complex diseases may be too minor to affect the discriminative accuracy of genomic profiling, and that given the small differences in effect sizes, the frequencies of the risk variants will determine the future feasibility of genomic profiling for the prediction of common diseases.

The results of the present studies demonstrate the importance of genotype frequency in the clinical validity of genomic profiling. Obviously, compared with genetic variants with weak effects, variants with strong effects lead to higher disease risks for those exposed. Yet variants with strong effects are generally rare and for that reason less likely to be useful for the prediction of common diseases because only a limited number of people will be exposed to multiple rare variants simultaneously. It has been demonstrated that the sensitivity and specificity of single genetic tests with two genotypes can only be maximal when the frequency of the risk variant is equal to the risk of disease[22]: predicting a rare disease by a common risk genotype or a common disease by a rare risk genotype by definition yields false-positive and false-negative findings, respectively. Similarly, maximal discriminative accuracy for genomic profiling—if possible at all—may only be realized when the combined frequency of what are considered at-risk profiles equals the disease risk.

Finally, the discriminative accuracy of type 2 diabetes prediction by genomic profiling using five susceptibility genes was poor (AUC = 0.58), but these results should be interpreted with caution. While effect estimates came from pooled studies or meta-analyses, these genetic variants are still subject to confirmation of their association with type 2 diabetes. The example of type 2 diabetes, in which only a limited number of susceptibility genes have been identified, is typical for common chronic diseases.[23] The unraveling of the genetic basis for chronic diseases is making remarkable progress and discoveries of new susceptibility genes and of gene-gene and gene-environment interactions are expected. This progress is facilitated by networks of investigators who combine their data to provide the large-scale epidemiological studies required for the identification of weak susceptibility genes (ORs <1.20) and interaction effects.[24,25] While it is expected that variants in TCF7L2 are likely the strongest genetic risk factors for type 2 diabetes,[26] this paper suggests that discoveries of common variants with weaker effects may still improve the discriminative accuracy of genomic profiling to a level useful for the prediction of disease in asymptomatic individuals. Nevertheless, several companies are prematurely selling online predictive genomic profiling for individualized nutrition and lifestyle recommendations based on a limited number of susceptibility genes.[27,28] The present analyses suggest that testing a small number of weak susceptibility genes (OR = 1.25) may not even be useful for the identification of at-risk individuals.

In conclusion, not only effect sizes of gene-disease associations but also genotype frequencies have a profound role in both the clinical validity of genomic profiling and the number of individuals who are at increased genetic risk. Both strong associations and frequent risk genotypes may benefit the feasibility of individualized medicine for common chronic diseases. Genotype frequencies should therefore be given greater attention when reporting the results of genetic association studies.

ACKNOWLEDGMENTS

The study was financially supported by the Center for Medical Systems Biology. Dr. Janssens was supported by a fellowship from The Netherlands Genomics Initiative.

References

1. Collins FS, McKusick VA. Implications of the Human Genome Project for medical science. *JAMA* 2001;285:540–544.
2. Valle D. Genetics, individuality, and medicine in the 21st century. *Am J Hum Genet* 2004;74:374–381.

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.

3. Evans A, Salomaa V, Kulathinal S et al. MORGAM (an international pooling of cardiovascular cohorts). *Int J Epidemiol* 2005;34:21–27.
4. Peltonen L. GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res* 2003;6:354–360.
5. Wilson PW, D'Agostino RB, Levy D, Belanger AM, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837–1847.
6. Chen P, Ratcliff G, Belle SH, Cauley JA, et al. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology* 2000; 55:1847–1853.
7. Yang Q, Khoury MJ, Botto L, Friedman JM, et al. Revisiting the clinical validity of multiplex genetic testing in complex diseases: reply to Janssens et al. *Am J Hum Genet* 2004;74:588–589.
8. Janssens ACJW, Aulchenko YS, Elefante S, Borsboom GJJM, et al. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med* 2006;8:395–400.
9. Yang Q, Khoury MJ, Friedman JM, Little J, et al. How many genes underlie the occurrence of common complex diseases in the population? *Int J Epidemiol* 2005.
10. Janssens ACJW, Pardo MC, Steyerberg EW, Van Duijn CM. Revisiting the clinical validity of multiplex genetic testing in complex diseases. *Am J Hum Genet* 2004;74: 585–588.
11. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology: a basic science for clinical medicine. Boston/Toronto: Little, Brown, 1985.
12. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002;31:33–36.
13. Peto J. Breast cancer susceptibility—a new look at an old model. *Cancer Cell* 2002; 1:411–412.
14. Harrell FE. Hmisc R function library, 2004. Available at: http://hesweb1.med.virginia.edu/biostat/s/library/r/. Accessed May 1, 2007.
15. Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol* 2006;164:609–614.
16. Narayan KM, Boyle JP, Thompson TJ, Sorensen SW, et al. Lifetime risk for diabetes mellitus in the United States. *JAMA* 2003;290:1884–1890.
17. Zeggini E, Parkinson JR, Halford S, et al. Examining the relationships between the Pro12Ala variant in PPARG and type 2 diabetes-related traits in UK samples. *Diabet Med* 2005;22:1696–1700.
18. van Dam RM, Hoebee B, Seidell JC, Schaap MM, et al. Common variants in the ATP-sensitive K^+ channel genes KCNJ11 (Kir6.2) and ABCC8 (SUR1) in relation to glucose intolerance: population-based studies and meta-analyses. *Diabet Med* 2005;22: 590–598.
19. Weedon MN, Schwarz PE, Horikawa Y, et al. Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am J Hum Genet* 2003;73:1208–1212.
20. Weedon MN, Owen KR, Shields B, et al. A large-scale association analysis of common variation of the HNF1alpha gene with type 2 diabetes in the U.K. Caucasian population. *Diabetes* 2005;54:2487–2491.
21. Zhang C, Qi L, Hunter DJ, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* 2006;55:2645–2648.
22. Khoury MJ, Newill CA, Chase GA. Epidemiologic evaluation of screening for risk factors: application to genetic screening. *Am J Public Health* 1985;75:1204–1208.
23. O'Rahilly S, Wareham NJ. Genetic variants and common diseases—better late than never. *N Engl J Med* 2006;355:306–308.
24. Ioannidis JP, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. *Am J Epidemiol* 2005;162:302–304.
25. Seminara D, Khoury MJO'Brien TR, et al. The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* 2007;18:1–8.
26. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881–885.
27. Sciona. Available at: http://www.sciona.com. Accessed August 23, 2006.
28. Genovations. Available at: http://www.genovations.com. Accessed August 23, 2006.

Copyright © American College of Medical Genetics. Unauthorized reproduction of this article is prohibited.