# ARTICLE

# Inclusion of Gene-Gene and Gene-Environment Interactions Unlikely to Dramatically Improve Risk Prediction for Complex Diseases

Hugues Aschard,[1,2,*] Jinbo Chen,[3] Marilyn C. Cornelis,[4] Lori B. Chibnik,[5] Elizabeth W. Karlson,[6] and Peter Kraft[1,2,7]

Genome-wide association studies have identified hundreds of common genetic variants associated with the risk of multifactorial diseases. However, their impact on discrimination and risk prediction is limited. It has been suggested that the identification of gene-gene (G-G) and gene-environment (G-E) interactions would improve disease prediction and facilitate prevention. We conducted a simulation study to explore the potential improvement in discrimination if G-G and G-E interactions exist and are known. We used three diseases (breast cancer, type 2 diabetes, and rheumatoid arthritis) as motivating examples. We show that the inclusion of G-G and G-E interaction effects in risk-prediction models is unlikely to dramatically improve the discrimination ability of these models.

## Introduction

Genome-wide association studies (GWASs) have discovered hundreds of common genetic variants associated with multifactorial diseases. These variants can be added to classical clinical and environmental risk factors for the improvement of risk-prediction assessment. However, for most common diseases, the addition of genetic variants to traditional risk factors has produced only modest improvements.[1–9] The subsequent genetic-risk profiles generated are still unlikely to provide sufficient discrimination to warrant individualized prevention except when the risks and costs associated with preventive interventions are low (but not so low that the benefits of intervention outweigh the costs for every individual).[10–12] It has been suggested that the identification of statistical interaction between genetic variants (G-G) or interaction between genetic variants and environmental risk factors (G-E) would improve the predictive accuracy of risk models and facilitate prevention.[2,12–14]

We explored the potential improvement in the discrimination of genetic-risk models for common complex diseases by using common markers and common exposures after incorporating G-G and G-E interactions. We considered the scenario in which multiple statistical interactions between common exposures and common risk markers exist and are known. We also allowed for the possibility that some genetic markers might be involved in G-G and G-E interactions yet have no marginal association with disease; hence, they would not have been detected with the standard one-marker-at-a-time GWAS approach. This is arguably a best-case scenario because few replicated interactions between common genetic markers and

other genetic and environmental factors have been identified to date;[15–18] also, after potential interactions are allowed for, very few novel markers that have undetectable marginal effect but that show evidence of association have been identified.[19]

Because the effectiveness of genetic-risk models will vary across diseases with different genetic architectures (defined by the number of associated loci and the distribution of the effect magnitude across loci), we considered three different hypothetical architectures motivated by breast cancer (BRCA, [MIM 114480]), type 2 diabetes (T2D, [MIM 125853]) and rheumatoid arthritis (RA, [MIM 180300]). We generated a broad range of hypothetical G-G and G-E interaction effects for each disease and constrained these models to be consistent with reported marginal effects for known genetic, clinical, and environmental risk factors. We then estimated the increase in discrimination ability of risk models that include the interactions and compared it to that of models that do not include the interactions.

## Material and Methods

### Simulation Scheme

We simulated hypothetical models in which the disease status was defined as a function of the marginal effects of known genetic and nongenetic risk factors (see Appendix A) and a random number of two-way G-G and G-E interaction effects. The interactions were defined such that the marginal effects of the known risk factors reflect the previously reported effect regardless of the number and the magnitude of the interactions that were simulated. We studied models in which statistical interactions exist only between the known risk factors as well as models in which

statistical interactions exist between the known risk factors and unknown risk SNPs that have no marginal effect. In each scenario, we generated 1,000 replicates of 100,000 women for whom the genotypes of the known SNPs and environmental risk were simulated independently, and we generated disease status by assuming a prevalence of 10% for BRCA, 9% for T2D, and 2% for RA.

The probability of disease given genotypes and exposures (penetrance) was generated with a constrained log-linear model:[20]

$$\log \Pr(D=1) = \min\left( A + \sum_j \beta_j G_j + \sum_l \sigma_l E_l + \sum_{j,k} I_{j,k} w_{j,k} \gamma_{j,k} + \sum_{l,m} I_{l,m} w_{l,m} \gamma_{l,m}, 0 \right).$$

(Equation 1)

Here, $\beta_j$ and $\sigma_l$ are the marginal effects of SNP $G_i$ and exposure $E_l$, respectively; $\gamma_{j,k}$ and $\gamma_{l,m}$ are the interaction effects between SNPs $j$ and $k$ and between the exposure $E_l$ and SNP $G_m$, respectively. $I_{jk}$ and $I_{lm}$ represent the presence (when equal to 1) or absence (when equal to 0) of interaction between $G_j$ and $G_k$ and between $E_l$ and $G_m$, respectively. $A$ is population-average log penetrance. $G_j$ is the weight for the genetic main effect, $E_i$ is the weight for the environmental main effect, $w_{j,k}$, is the weight of the interaction between SNPs $j$ and $k$, and $w_{l,m}$ is the weight of the interaction between exposure $E_l$ and SNP $G_m$ (Appendix B). These weights are chosen so that the marginal effect of SNP $j$—that is, the difference in the log penetrance averaged over other genes and exposures between subjects carrying one versus zero (or two versus one) copies of the minor allele for SNP $j$—equals $\beta_j$ and so that the marginal effect of exposure $l$ is $\sigma_l$. In particular, the expected effect of each interaction term is equal to 0 within strata defined by either of the following interacting factors:

$$E(w_{j,k}\gamma_{j,k} \mid G_k) = E(w_{j,k}\gamma_{j,k} \mid G_j) = 0 \text{ and } E(w_{l,m}\gamma_{l,m} \mid E_l)$$
$$= E(w_{l,m}\gamma_{l,m} \mid G_m) = 0.$$

This model is identical to the model generated by the commonly used parameterization

$$\log \Pr(D=1) = \alpha + \sum_j \beta_j^* G_j^* + \sum_l \sigma_l^* E_l^* + \sum_{j,k} I_{j,k} G_j^* G_k^* \gamma_{j,k} + \sum_{l,m} I_{l,m} E_l^* G_m^* \gamma_{l,m},$$

(Equation 2)

where $G_j^*$ is the count of minor alleles for SNP $j$ and $E_l^*$ is an indicator variable for whether a subject is exposed (Appendix B). Moreover, the interaction parameters $\gamma_{*,*}$ in Equations 1 and 2 are equivalent and have the same interpretation. We chose Equation 1 over Equation 2 because of the equivalence between the marginal effects and the parameters $\beta_j$ and $\sigma_l$. This allowed us to easily constrain the simulated models to have locus-specific marginal effects and exposure marginal effects similar to those observed empirically. The simulated models cover a wide range

of interaction patterns, including classical supermultiplicative and submultiplicative effects and more exotic models close to those described by Evans et al.[21]

We varied the number of pairwise interactions from one to five G-G and one to five G-E effects. The interaction effects $\gamma_{j,k}$ and $\gamma_{l,m}$ were sampled from a normal distribution such that $|\gamma| < \ln(2)$ with 95% probability. We chose the range of interaction odds ratios to be consistent with (1) reported G-E interaction odds ratios, which are mostly smaller than 2.5;[22–27] (2) the lack of consistently replicated large G-G and G-E interactions involving common variants; and (3) the magnitude of the marginal effects seen in GWASs. Depending on the scenario we considered, the SNPs noted with subscripts $k$ and $m$ represent either some of the known risk SNPs or some hypothetical unidentified risk SNPs that have no marginal effect.

(We recognize that some of the "environmental" risk factors that we have chosen—e.g. body mass index (BMI) or family history of disease—are endogenous factors that might themselves be under genetic influence or correlated with the risk alleles. Our simulation model assumes that the risk alleles are not correlated with the environmental risk factors and that any loci influencing the "environmental" risk factors have no unmediated influence on disease risk. This would be the case for many exogenous exposures that are not correlated with genotypes at the disease susceptibility loci or for loci whose effect is solely mediated through the endogenous environment, as might be the case for FTO, BMI, and T2D. We did not explore situations in which loci influencing the "environmental" risk factors also have an unmediated influence on disease given that for most complex diseases, the overlap between currently known risk loci and loci associated with "environmental" risk factors is small. We stress that our simulations focus on the best-case scenario in which the penetrance model is known; in particular, markers that have no effect on disease risk conditional on the "environmental" exposure are not included in the penetrance model, and all relevant endogenous exposures are included in the penetrance model.).

## Estimation of Improvement in Risk Prediction

The improvement in discrimination ability was calculated as the difference in the area under the receiver operating characteristic curve (AUC), also called the C-statistic, between the model that includes only the marginal effect of the known risk factors and the model that includes the main effects of risk factors and the simulated interaction effects. The improvement in estimation of absolute risk was assessed by the continuous net reclassification index (cNRI)[28] before and after the addition of the interaction effects. The cNRI quantifies the proportion of individuals who have improved absolute-risk estimates (i.e., cases that tend to have a higher absolute risk and controls that tend to have a smaller absolute risk), although the cNRI does not consider the magnitude of the changes in risk estimates. Unlike the net reclassification index (NRI),[29] the cNRI does not require the specification of risk categories and thus allows for direct comparison across different diseases. It is defined as follows:

$$\text{cNRI} = \frac{\Pr(\text{event}|\text{up}) \times \Pr(\text{up}) - \Pr(\text{event}|\text{down}) \times \Pr(\text{down})}{P(\text{event})} + \frac{\Pr(\text{nonevent}|\text{down}) \times \Pr(\text{down}) - \Pr(\text{nonevent}|\text{up}) \times \Pr(\text{up})}{P(\text{nonevent})},$$

**Table 1. Expected Discrimination Ability for BRCA, RA, and T2D Based on Reported Estimated Effect of Genetic and Nongenetic Risk Factors**

| Disease | AUC (SD) | | | Sensitivity by Specificity[a] (SD) | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Genetic Model | Nongenetic Model | Combined Model | Genetic Model | Nongenetic Model | Combined Model |
| BRCA | 0.591 (0.008) | 0.558 (0.008) | 0.616 (0.010) | 0.023 (0.005) | 0.017 (0.004) | 0.033 (0.010) |
| RA[b] | 0.687 (0.007) | 0.571 (0.009) | 0.704 (0.008) | 0.055 (0.007) | 0.014 (0.004) | 0.066 (0.009) |
| T2D | 0.623 (0.009) | 0.754 (0.007) | 0.786 (0.006) | 0.030 (0.005) | 0.088 (0.007) | 0.122 (0.012) |

The following abbreviations are used: AUC, area under the receiver operating characteristic curve; SD, standard deviation; BRCA, breast cancer; RA, rheumatoid arthritis; and T2D, type 2 diabetes.
[a]For a specificity threshold of 0.99.
[b]RA genetic model and combined model include both SNPs and HLA variants.

where Pr(event) and Pr(nonevent) are the probabilities of being a case or a control, respectively, and Pr(down) and Pr(up) are the probabilities of having lower or higher risk estimates, respectively, after the addition of interaction effects in the model. The cNRI and AUC are independent of disease prevalence, allowing these measures to be directly compared across diseases.

## Results

### Disease Models and Marginal Risk Prediction
For the BRCA risk model, we included 15 common SNPs with marginal relative risks (RRs) ranging from 1.07 to 1.28 and risk-allele frequencies ranging from 0.15 to 0.85 (Table S1, available online). Four of the five components of the Gail model [30] were added; these were age at menarche, the number of previous breast biopsies, age at first child, and the number of first-degree relatives who have developed BRCA (age < 50 was not taken into account because we considered a situation in which all women were older than 50). We modeled the joint frequencies of these risk factors by using their empirical distribution in the Nurses' Health Study (NHS) (see Material and Methods and Table S2). For T2D, we included 31 SNPs and four clinical risk factors. The marginal allelic RRs of the SNPs ranged from 1.09 to 1.40, and risk-allele frequencies were between 0.10 and 0.93 (Table S3). The clinical risk factors were obesity (BMI ≥ 30), smoking status, physical activity, and family history of T2D. We used the NHS to model the joint frequency of these factors (Table S4). With a RR of 5.1 and a frequency of 0.25, obesity was the strongest risk factor of T2D. The three others have RRs between 1.11 and 3.03 and frequencies between 0.16 and 0.56. The RA model included 31 SNPs with RRs ranging from 1.02 to 1.75 and risk-allele frequencies ranging from 0.08 to 0.90 (Table S5). Other than the 31 common GWAS-identified SNPs, we included eight *HLA-DRB* alleles. The largest effect was for *DRB\*0401*, which has a frequency of 0.09 for a RR of 3.30, whereas the other variants had either a low frequency or a low RR. For ease of computation, all HLA risk alleles were merged into a single variant that had an average RR of 2.33 for a frequency of 0.24. We included only two nongenetic risk factors, smoking and breast feeding, which have RRs equal to 1.69 and 0.79, respec-

tively. The frequencies of these risk factors were also taken from the NHS (Table S6).

For each disease, we generated 1,000 replicates of 100,000 women and assigned them a disease status on the basis of the effects described above. To provide a benchmark for later simulations, we initially assumed that there were no G-G or G-E interactions. The smallest discrimination ability as estimated by our simulation was observed for the BRCA model: the average AUC for the model that combined genetic and nongenetic factors was 0.616 (standard deviation [SD] = 0.010). Conversely, T2D had the highest discrimination ability—it had an AUC of 0.786 (SD = 0.006) for the combined model. The AUC for RA was 0.704 (SD = 0.008) and was driven by its genetic component, which was proportionally much higher than that of the two other models. Overall, the AUCs of the genetic, nongenetic, and combined models obtained by simulation (Table 1) were in agreement with the AUCs reported on real data in the literature (see Appendix B).

### Increase in Discrimination if G-G and G-E Interactions Exist between Known Risk Factors
We added between two and ten G-G and G-E interaction effects to the simulation models so that the final marginal effects of known genetic risk variants and clinical and environmental risk factors were equal to their previously reported effects (see Material and Methods and Appendix B). We compared the models that included the simulated interactions with models that included only the marginal effects (referred to hereafter as "marginal models") in terms of their ability to discriminate and appropriately reclassify cases and controls.

The addition of interaction effects to the risk model improved the performance of genetic-risk-prediction models, although the improvement as measured by AUC and cNRI was modest. (Because we simulated a large cohort, the improvement in model fit was almost always statistically significant [the median likelihood-ratio p value $\approx 10^{-23}$] whether measured by the differences in the log likelihood or changes in AUC or cNRI. Our focus is on the magnitude of these differences.) The cNRI increased linearly with the number (Figure 1) and the magnitude (Figure 2) of the simulated interactions. The average cNRI over all simulations was similar across all
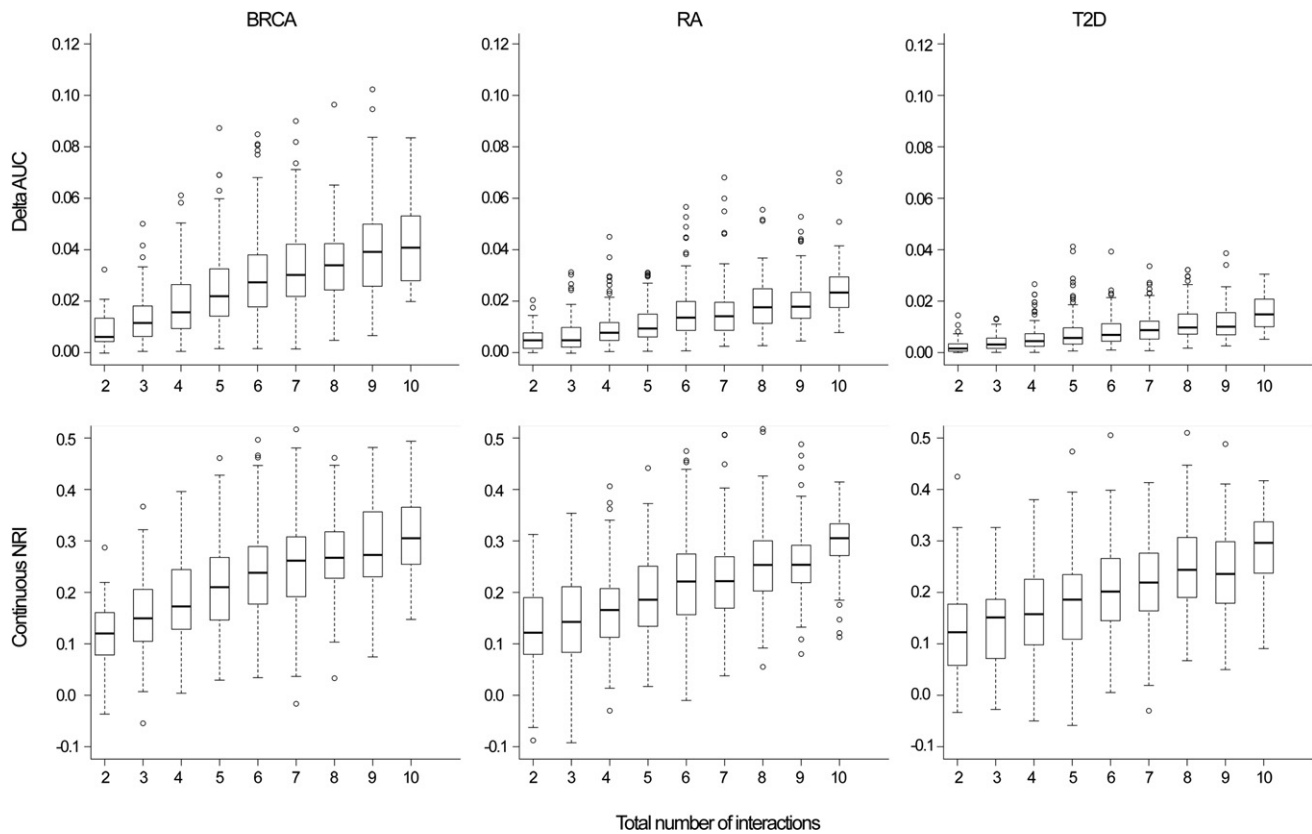
**Figure 1. Absolute Increase in AUC and cNRI by the Number of Interactions Simulated**
Breast cancer (BRCA), rheumatoid arthritis (RA), and type 2 diabetes (T2D).

three disease models: 0.230 (SD = 0.096), 0.210 (SD = 0.092), and 0.200 (SD = 0.096) for BRCA, RA, and T2D, respectively. The contribution of controls to the cNRI was slightly higher than the contribution of cases. Approximately 54% of cases had higher risk estimates after interactions were incorporated, whereas 56% of controls had lower estimates. The highest cNRI was observed when ten interactions were simulated (0.312 [SD = 0.083], 0.305 [SD = 0.085], and 0.286 [SD = 0.074] for BRCA, RA, and T2D, respectively); the proportions of cases and controls that had improved risk estimates in this situation were 57% and 59%, respectively.

The improvement in discrimination ability as measured by the AUC differed by risk model (Figures 1 and 2). The average absolute increase in AUC was 2.82% (SD = 1.74), 1.40% (SD = 1.00), and 0.85% (SD = 0.63) for BRCA, RA, and T2D, respectively. These differences reflect the increasing difficulty in producing the same absolute change in AUC as the baseline AUC increases. The increase in the difference in mean predicted risk between cases and controls after the addition of interactions was similar for all three disease models; however, the variance of predicted values of cases and controls was higher for models with a higher baseline AUC, resulting in a lower increase in the AUC of these models.[31]

We limited the magnitude of the interaction in the simulations presented in Figures 1, 2, and 4 to relatively small effects ($0.5 < RR_{interaction} < 2$ with 95% chances) given that few larger interaction effects involving known risk factors for complex diseases have been detected by large-scale studies conducted on humans to date. In Figure S1, we explore the impact of larger interaction effects (relative risk up to 10) and a larger number of interactions (up to 20, including interactions with SNPs that have no marginal effect). The improvement in AUC increased with increasing interaction effect; the magnitude of this increase again differed by disease. For example, for 4–20 interactions, the average absolute improvement in AUC for $0.2 < RR_{interaction} < 5$ was equal to 13.79% (SD = 4.64), 8.21% (SD = 3.19), and 5.35% (SD = 2.37) for BRCA, RA, and T2D, respectively.

### Increase in Discrimination if Interactions Exist between Known Risk Factors and SNPs with No Marginal Effect

We estimated the improvement when interactions between known risk factors and SNPs that have no marginal effect might also exist. We reconducted the same simulations but replaced the known SNPs with SNPs that we generated to have risk-allele frequencies in the interval [0.05, 0.95] and no marginal effect on the diseases studied. These simulations show similar results to the previous scenario, but the improvement was, on average, slightly lower when compared with the model that
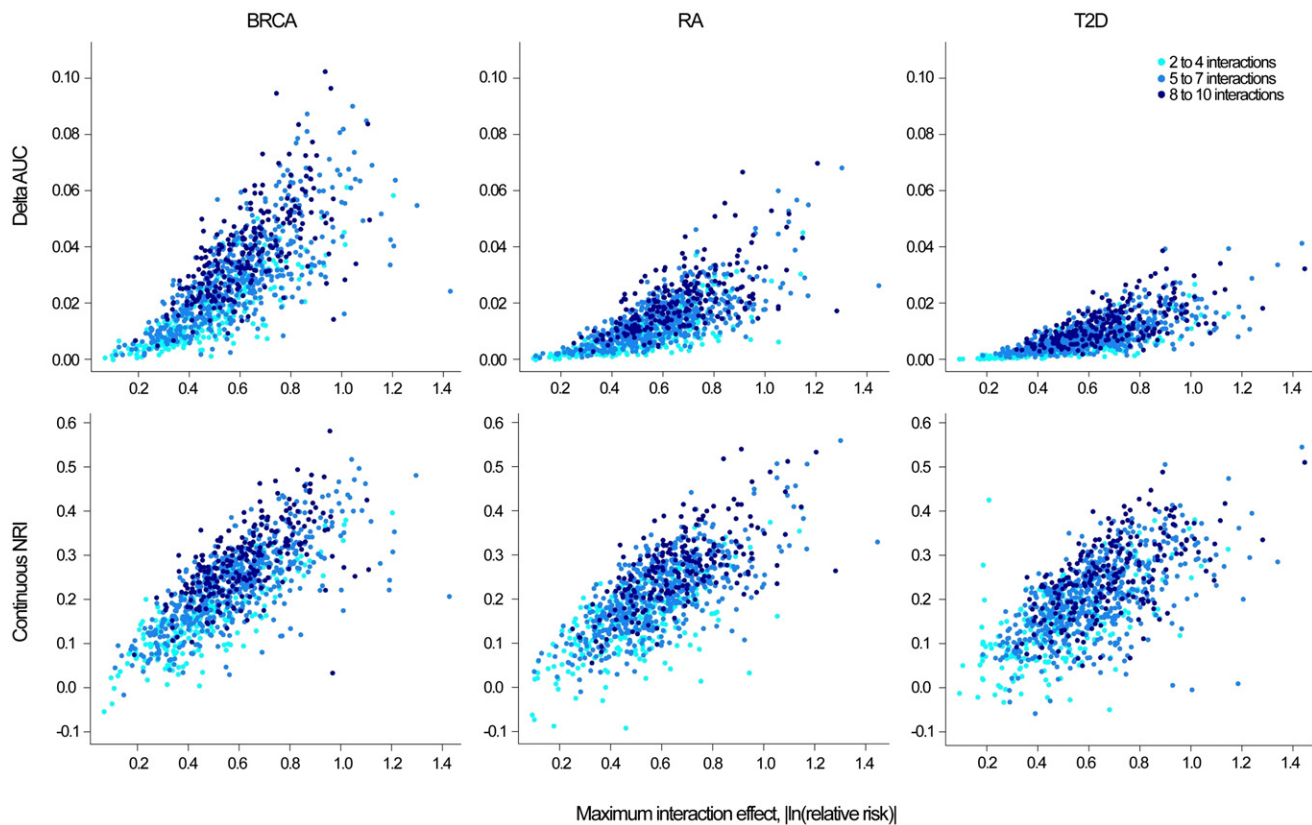
**Figure 2.   Absolute Increase in AUC and cNRI by the Maximum Interaction Effect Simulated**
Breast cancer (BRCA), rheumatoid arthritis (RA), and type 2 diabetes (T2D).

included interaction between known risk factors only. The average absolute improvement in AUC was equal to 2.32% (SD = 1.46), 1.10% (SD = 0.69), and 0.76% (SD = 0.56) for BRCA, RA, and T2D, respectively; when interactions only involved known risk SNPs, the average absolute improvement in AUC was 2.82%, 1.40%, and 0.85% for BRCA, RA, and T2D, respectively.

### Number of Interactions versus Size of the Interaction

We further investigated whether model improvement was more sensitive to the size of the interaction effects (as measured by the log-interaction RR denoted as $\gamma$) or to the number of interactions. We compared two additional models. In the first, we fixed the number of interactions at ten and limited the size of the interactions ($|\gamma| <$ ln(1.2) with 95% probability); in the second, we fixed the number of interactions at two and simulated only medium-to-strong interactions (ln(1.5) $< |\gamma| <$ ln(2) with 95% probability). The comparison of the increase in AUC between these two models for each of the diseases when 1,000 series of each model were simulated is presented in Figure 3. For all models, the magnitude of interaction leads to a slightly better AUC improvement than does the number of interactions. As before, the increase in AUC decreased as the discrimination of the marginal model increased, and the highest difference was observed for BRCA. For this model, the absolute increase in AUC was

on average 2.32% (SD = 0.93) for two strong interactions and 0.63% (SD = 0.30) for ten low interactions. For RA, the increases for two and ten interactions were, respectively, 1.00% (SD = 0.44) and 0.64% (SD = 0.31); for T2D, the increases were 0.27% (SD = 0.14) and 0.15% (SD = 0.08), respectively.

### Impact on Public Health and Clinical Utility

Interpreting the AUC and cNRI can be difficult in terms of public-health impact because they do not consider different benefits and costs associated with true, false-positive, and negative results.[32] Other measures can be used depending on context.[33–35] Here, we consider sensitivity for a fixed specificity threshold, which corresponds to a single point on the receiver operating characteristic (ROC) curve. This measure is relevant when a driving consideration is the control of false-positive test results, such as in the case of a population-screening program in which the intervention that follows a positive test for a rare disease carries substantial risk. The sensitivity increase due to the inclusion of interactions for different specificity thresholds is presented in Figure 4. The average absolute increase in sensitivity (the percentage of cases above the risk threshold) was below 0.01 for all disease models at a high level of specificity ($\geq 0.99$), whereas for lower specificity thresholds ($\geq 0.90$), the improvement was moderate: 0.033 (SD = 0.001), 0.019 (SD = 0.000),
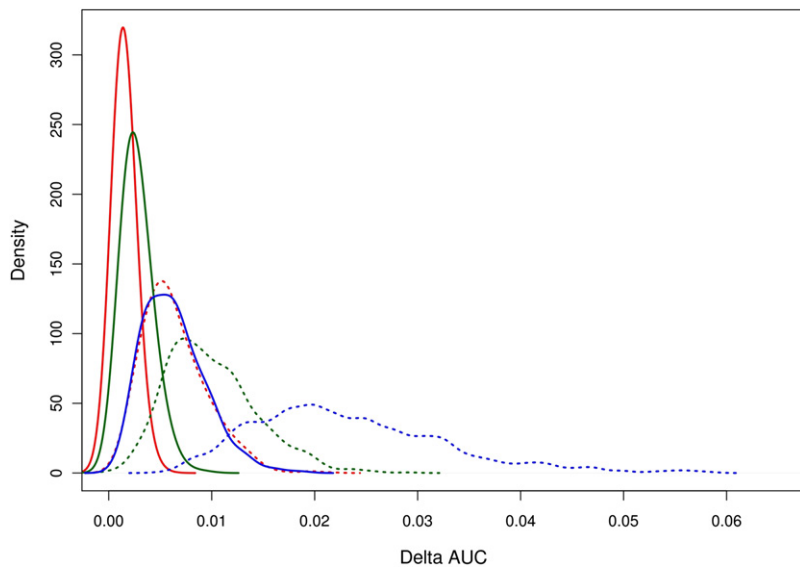
and 0.014 (SD = 0.000) for BRCA, RA, and T2D, respectively. This suggests that there might be modest sensitivity gains for intermediate risk thresholds with higher false-positive proportions (>10%), but the gains for high risk thresholds with small false-positive proportions (<1%) might be low.

## Discussion

Although multiple biologic interactions among GWAS-identified risk loci and clinical risk factors are likely to contribute to the etiology of many common diseases, this study suggests that the identification of statistical interactions among these factors might have a modest impact on risk prediction and discrimination for common complex diseases. Moreover, the improvement in risk prediction was estimated in this study in a best-case scenario in which the true main effect and interaction effect are known and are homogeneous across all individuals (this might not be the case in practice). The discriminatory and predictive accuracy of any model trained in a finite (and perhaps relatively small) dataset will be diminished as a result of model-selection uncertainty (e.g., incorrectly excluding or including true risk factors or interactions) and variability in parameter estimates. Our quantitative results are in agreement with the claim made by Gail[2] that G-G interactions might have a low effect on discrimination accuracy.

The evidence to date suggests that strong pairwise-interaction effects (e.g., RR > 2) between known risk SNPs and classical clinical risk factors are unlikely to be common for complex diseases. For example, three large studies of interactions among GWAS-identified BRCA markers and established risk factors failed to find any compelling evidence for interaction despite having greater than 90% power for the detection of interaction RR as low as 1.06,[15,18,36] and there are few validated interactions among established genetic or dietary risk factors for T2D.[37] We chose the distribution of the magnitude of simulated interaction effects to be consistent with these equivocal results.

The paucity of replicated statistical G-G and G-E interactions from observational studies in humans contrasts with
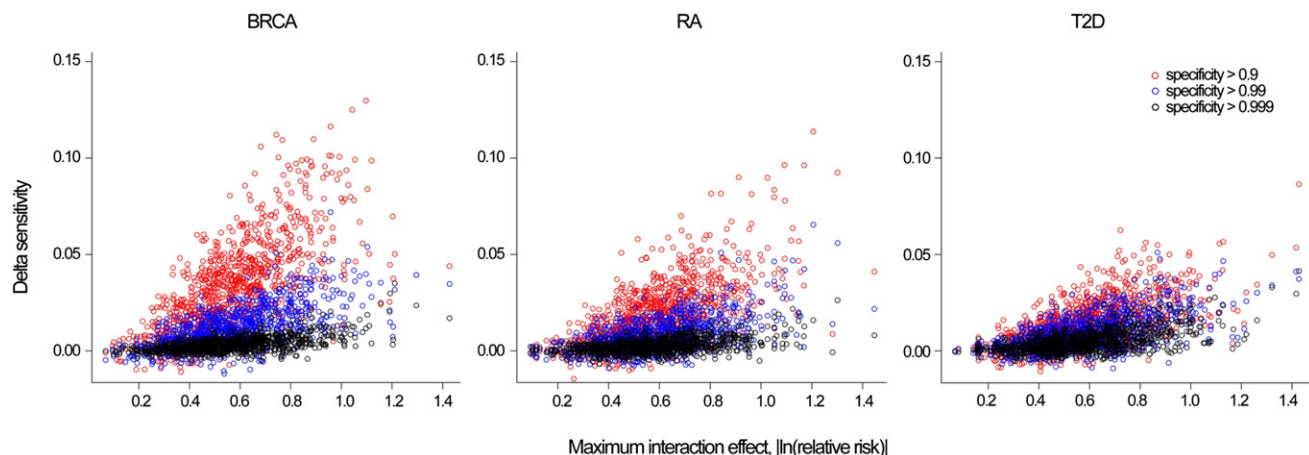


**Figure 4. Absolute Increase in Sensitivity by Maximum Interaction Effect and Specify Threshold**
Breast cancer (BRCA), rheumatoid arthritis (RA), and type 2 diabetes (T2D).

the rich literature on these interactions in experimental and free-living model organisms.[38,39] A number of factors might contribute to this contrast, and these include (but are not limited to) low power for the detection of statistical interactions as a result of the tagging-SNP approach adopted by GWASs and many candidate-gene studies,[40] limited genetic and environmental diversity in the human studies to date,[41,42] or the actual absence of detectable statistical interactions in many human populations. Our extended simulations show that the inclusion of many interactions (>10) with modest to large effects (interaction odds ratios >2) might meaningfully increase model discrimination. It remains an open empirical question whether such interactions exist.

Large pairwise or higher-order interaction effects involving rare causal variants, rare exposures, or rare allelic combinations of common risk variants might exist but were outside of the scope of this study. Pharmacogenetics is one setting in which large interactions have been observed.[43,44] The odds ratio for adverse drug reactions for some variants can be relatively large (well over 5), and, considering that the adverse reaction is typically not present among those who do not receive the drug, the gene-drug interaction odds ratio can be thought of as infinite. However, our focus was on risk screening in the general population, not on tailoring therapy.

Large interactions involving rare exposures or rare alleles might help identify individuals who are at particularly high risk, but because they are rare and the prior probability for particular interactions will often be low, reliably identifying them will be difficult. The impact of a strategy that focuses on high-risk individuals might offer substantial benefits for those individuals. However, because most cases arise among the many at lower risk rather than among the few who are at high risk, the potential impact on the total burden of disease in the population might be limited.[45]

There are many measures that can be used for summarizing the clinical and public-health utility of risk-prediction models.[33,46] We used the AUC. Other measures such as risk-reclassification tables or the NRI can be used;[29] however, a recent study by Mihaescu et al. has shown that these reclassification measures increase as the AUC increases.[47] Most reclassification measurements depend on risk thresholds that place individuals into risk categories; different thresholds yield different measures and can change the relative ranking of risk-prediction models. Consequently, we used the cNRI, which is a threshold-free reclassification tool. The cNRI simply measures the proportion of individuals who have better risk estimates (cases whose predicted risk increases after the inclusion of interactions and controls whose predicted risk decreases) without quantifying the magnitude of the change in risk estimates. Finally, we also considered the increase in sensitivity for high-specificity thresholds, a more interpretable measure for public-health recommendation. The inclusion of G-G and G-E interactions did not dramatically improve any of these measures.

The identification of G-G and G-E interactions remains of major interest because it can provide important clues regarding the biological mechanism of many common complex diseases.[48–50] However, indentifying replicable statistical interactions between common genetic and non-genetic risk factors is a very challenging task. Such interactions might only modestly improve risk models designed for use in the general population (such the Gail score or the Framingham risk score). Whether such modest improvements are clinically important will depend on context. Our results suggest that the improvement of risk-prediction models (as measured by the change in average sensitivity) with the use of risk factors with modest effects primarily depends on the number of risk factors included in the model; including nonlinear terms generally provides little improvement. This suggests that identifying new risk loci and new environmental risk factors might lead to greater improvements in risk modeling than incorporating interactions among known risk factors. We note, however, that G-G and G-E interactions can be leveraged for the identification of new risk loci and new modifiable environmental risk factors.[19,50–54] Finally, we stress that although G-G and G-E interactions might have modest impacts on risk prediction, an understanding of the interplay between genes and the environment can provide insights into disease etiology; this understanding, in turn, can lead to improved treatment and prevention strategies.

## Appendix A

The BRCA risk model included SNPs that GWASs have found to be associated with BRCA at highly stringent levels of statistical significance in populations of European ancestry and that were confirmed in at least one independent set of cases and controls (Table S1). On the basis of the catalog of published GWASs,[55] we identified 12 SNPs that match these criteria at the time we were conducting this study. We also considered three additional SNPs that match similar criteria but that have been identified in post-GWAS analysis involving multiple independent populations. We added these 15 SNPs to the model that we simulated by using their previously reported estimated marginal effects. We simulated these SNPs by using the frequencies obtained from the HapMap CEU[56] (Utah residents with northern and Western European ancestry from the CEPH collection) samples and the reported per-allele RRs from discovery studies. We combined these SNPs with the four components of the Gail score for women older than 50; these components were age at menarche, age at first child, number of first-degree relatives with BRCA, and the number of previous breast biopsies (Table S2). The odd ratios of these risk factors were extracted from Gail et al.[30] The frequencies were taken from 1,142 control samples from a BRCA case-control study nested within the NHS.[57]

**Table A1. Coding for Terms _G_, _E_, and _w_ from Equation 1 and Terms _G*_ and _E*_ from Equation 2**

| Stratum | G* | E* | G | E | w | Frequency |
|---|---|---|---|---|---|---|
| Noncarrier, unexposed | 0 | 0 | $-q$ | $-\pi$ | $q\pi$ | $(1-q)(1-\pi)$ |
| Carrier, unexposed | 1 | 0 | $(1-q)$ | $-\pi$ | $-(1-q)\pi$ | $q(1-\pi)$ |
| Noncarrier, exposed | 0 | 1 | $-q$ | $(1-\pi)$ | $-q(1-\pi)$ | $(1-\pi)q$ |
| Carrier, exposed | 1 | 1 | $(1-q)$ | $(1-\pi)$ | $(1-q)(1-\pi)$ | $q\pi$ |

Parameters $q$ and $\pi$ are the minor allele frequency of the SNP $G$ and the frequency of the exposure $E$, respectively.

The T2D risk model included loci that have been found to be associated with T2D at highly stringent levels of statistical significance; these loci were identified or replicated by the Diabetes Genetics Replication and Meta-Analysis (DIAGRAM) consortium, a widespread collaboration that studies populations of European descent (the most recent report includes 42,542 case subjects with T2D and 98,912 control subjects). We used the most significant SNPs of each loci from the DIAGRAM consortium and the corresponding reported estimated effects, and we used frequencies extracted from the HapMap CEU[56] (Table S3) samples. We considered four established clinical risk factors: obesity (BMI ≥ 25), smoking status, physical activity, and family history of T2D (Table S4). Given that no established risk score exists for T2D, estimated effects of these risk factors were extracted from the NHS. Their joint frequencies were taken from control samples in a case-control study of T2D in the NHS.[58]

The RA risk model included established risk factors of seropositive RA. Selected genetic variants include eight _HLA-DRB1_ alleles, SNPs newly identified in GWASs, and SNPs from candidate genes that have been confirmed in GWASs to be at a genome-wide significance level. The estimated effects were derived from the largest GWASs and a recent meta-analysis of all published studies[59] (Table S5). We simulated these SNPs by using the HapMap CEU[56] frequencies, and we simulated the DRB loci by using frequencies extracted from a sample of NHS controls.[60]

We considered two established clinical risk factors—breast feeding and smoking status (Table S6). The frequencies of these risk factors were extracted from the NHS case-control study of RA,[61] whereas the estimated effects were extracted from the discovery studies.[62,63]

## Appendix B

To evaluate the relevance of the simulated models, we estimated, as an example, the discrimination ability of the marginal model used for BRCA in a real dataset. For this aim, we used 1,139 BRCA cases and 1,140 controls from the NHS study.[64] All women considered were post-menopausal and older than 50. The 15 SNPs of interest were either genotyped as part of the Cancer Genetic Markers of Susceptibility Study (CGEMS)[65] with the Illumina 550K or imputed with the software MACH[66] and the HapMap (rel22) CEU data.[56] The four Gail components were extracted from the detailed NHS follow-up questionnaires that have been mailed biennially to the full NHS cohort for updating exposure information and any major medical events.

The genetic model including the 15 SNPs yielded an AUC of 0.597 (standard error of the mean [SEM] = 0.012), the nongenetic model including the four Gail component yielded an AUC of 0.583 (SEM = 012), and the combined Gail and SNP model shows an increase in prediction ability and yielded an AUC of 0.629 (SEM = 0.012). These results are in agreement with a previous study that showed similar results for the Gail model and slightly lower results for the Gail and SNP models when using a lower number of SNPs (AUC = 0.618 when Gail and ten SNPs were used in Wacholder et al.;[4] AUC = 0.594 when Gail and seven SNPs were used in Mealiffe et al.[3]). In comparison, our simulated model with the same estimates yielded an average AUC of 0.558 (SD = 0.008), 0.591 (SD = 0.008), and 0.616 (SD = 0.010) for the Gail model, the genetic model, and the combined model, respectively, when no interaction was simulated.

We simulated the interaction effects while constraining the model to have locus-specific and exposure marginal effects similar to those observed empirically. Rather than using the popular parameterization that is based on

**Table A2. Equivalence between Penetrance Models from Equations 1 and 2**

| Stratum | Penetrance According to Equation 1 | | According to Equation 2 |
|---|---|---|---|
| Noncarrier, unexposed | $A - \beta q - \sigma\pi + \gamma q\pi$ | = | $\alpha$ |
| Carrier, unexposed | $A + \beta(1-q) - \sigma\pi - \gamma(1-q)\pi$ | = | $\alpha + \beta\star$ |
| Noncarrier, exposed | $A - \beta q + \sigma(1-\pi) - \gamma q(1-\pi)$ | = | $\alpha + \sigma\star$ |
| Carrier, exposed | $A + \beta(1-q) + \sigma(1-\pi) + \gamma(1-q)(1-\pi)$ | = | $\alpha + \beta\star + \sigma\star + \gamma\star$ |

Parameters $q$ and $\pi$ are the minor allele frequency of the SNP $G$ and the frequency of the exposure $E$, respectively.

counts of minor alleles for SNPs and indicator variables for exposures and that includes product interaction terms (Equation 2), we used a coding scheme that ensures that the SNP and exposure main-effect parameters $\beta_j$ and $\sigma_l$ are equivalent to the desired marginal effects (Equation 1). We stress, however, that the two parameterizations yield identical models and, moreover, that the interaction terms $\gamma_{*,*}$ are equivalent.

We illustrate these two parameterizations in the simple case in which there is one haploid SNP and one binary exposure. (The results extend naturally to multiple diploid SNPs and multiple exposures when it is assumed that the population is in Hardy-Weinberg equilibrium, the SNPs are unlinked, and there is no correlation between the SNPs and exposures.) Table A1 presents the coding for terms $G$, $E$, and $w$ from Equation 1 and terms $G^*$ and $E^*$ from Equation 2 over the four possible genotype-exposure combinations along with each combination's frequency derived with the SNP minor allele frequency $q$ and the exposure frequency $\pi$. These codings ensure that the expectations for $E$ and $w$ are 0 within genotype strata and that the expectations for $G$ and $w$ are also 0 within exposure strata. From this and Equation 1, it follows that the difference in average log penetrance between carriers and noncarriers of the minor allele is $\beta$ and that the difference in average log penetrance between exposed and unexposed subjects is $\sigma$. The equivalence between the interaction parameter $\gamma$ in both models can be seen by the solution of the system of equations described in Table A2. This yields $\beta^* = \beta - \gamma\pi$, $\sigma^* = \sigma - \gamma q$, and $\gamma^* = \gamma$.

We validated our simulation model by comparing the estimated marginal effects of the known risk factors while either simulating or not simulating interactions. As an indicator of the goodness of fit, the mean of the average absolute distance between the estimated marginal additive effects and the true marginal additive effects for the BRCA model was equal to 0.030 (0.024 and 0.044 for the SNPs and the four Gail covariates, respectively) when G-G and G-E interactions between known risk factors were simulated.

## Supplemental Data

Supplemental Data include one figure and six tables and can be found with this article online at http://www.cell.com/AJHG.

## Web Resources

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

## References

1. Janssens, A.C., and van Duijn, C.M. (2008). Genome-based prediction of common diseases: Advances and prospects. Hum. Mol. Genet. *17* (R2), R166–R173.
2. Gail, M.H. (2008). Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J. Natl. Cancer Inst. *100*, 1037–1041.
3. Mealiffe, M.E., Stokowski, R.P., Rhees, B.K., Prentice, R.L., Pettinger, M., and Hinds, D.A. (2010). Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. J. Natl. Cancer Inst. *102*, 1618–1627.
4. Wacholder, S., Hartge, P., Prentice, R., Garcia-Closas, M., Feigelson, H.S., Diver, W.R., Thun, M.J., Cox, D.G., Hankinson, S.E., Kraft, P., et al. (2010). Performance of common genetic variants in breast-cancer risk models. N. Engl. J. Med. *362*, 986–993.
5. Cornelis, M.C., Qi, L., Zhang, C., Kraft, P., Manson, J., Cai, T., Hunter, D.J., and Hu, F.B. (2009). Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann. Intern. Med. *150*, 541–550.
6. Talmud, P.J., Hingorani, A.D., Cooper, J.A., Marmot, M.G., Brunner, E.J., Kumari, M., Kivimäki, M., and Humphries, S.E. (2010). Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ *340*, b4838.
7. Paynter, N.P., Chasman, D.I., Paré, G., Buring, J.E., Cook, N.R., Miletich, J.P., and Ridker, P.M. (2010). Association between a literature-based genetic risk score and cardiovascular events in women. JAMA *303*, 631–637.
8. Spitz, M.R., Amos, C.I., D'Amelio, A., Jr., Dong, Q., and Etzel, C. (2009). Re: Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J. Natl. Cancer Inst. *101*, 1731–1732, author reply 1732.
9. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. *40*, 310–315.
10. Haga, S.B., Khoury, M.J., and Burke, W. (2003). Genomic profiling to promote a healthy lifestyle: Not ready for prime time. Nat. Genet. *34*, 347–350.
11. Rose, G. (1985). Sick individuals and sick populations. Int. J. Epidemiol. *14*, 32–38.
12. Khoury, M.J., Yang, Q., Gwinn, M., Little, J., and Dana Flanders, W. (2004). An epidemiologic assessment of genomic profiling for measuring susceptibility to common diseases and targeting interventions. Genet. Med. *6*, 38–47.
13. Pharoah, P.D., Antoniou, A.C., Easton, D.F., and Ponder, B.A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. N. Engl. J. Med. *358*, 2796–2803.
14. Moore, J.H., and Williams, S.M. (2009). Epistasis and its implications for personal genetics. Am. J. Hum. Genet. *85*, 309–320.
15. Campa, D., Kaaks, R., Le Marchand, L., Haiman, C.A., Travis, R.C., Berg, C.D., Buring, J.E., Chanock, S.J., Diver, W.R.,

Dostal, L., et al. (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J. Natl. Cancer Inst. *103*, 1252–1263.

16. Lindstrom, S., Schumacher, F., Siddiq, A., Travis, R.C., Campa, D., Berndt, S.I., Diver, W.R., Severi, G., Allen, N., Andriole, G., et al. (2011). Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers—results from BPC3. PLoS ONE *6*, e17142.

17. Padyukov, L., Silva, C., Stolt, P., Alfredsson, L., and Klareskog, L. (2004). A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum. *50*, 3085–3092.

18. Travis, R.C., Reeves, G.K., Green, J., Bull, D., Tipper, S.J., Baker, K., Beral, V., Peto, R., Bell, J., Zelenika, D., and Lathrop, M.; Million Women Study Collaborators. (2010). Gene-environment interactions in 7610 women with breast cancer: Prospective evidence from the Million Women Study. Lancet *375*, 2143–2151.

19. Hamza, T.H., Chen, H., Hill-Burns, E.M., Rhodes, S.L., Montimurro, J., Kay, D.M., Tenesa, A., Kusel, V.I., Sheehan, P., Eaaswarkhanth, M., et al. (2011). Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. PLoS Genet. *7*, e1002237.

20. Wray, N.R., and Goddard, M.E. (2010). Multi-locus models of genetic risk of disease. Genome Med. *2*, 10.

21. Evans, D.M., Marchini, J., Morris, A.P., and Cardon, L.R. (2006). Two-stage two-locus models in genome-wide association. PLoS Genet. *2*, e157.

22. Lewis, S.J., and Smith, G.D. (2005). Alcohol, ALDH2, and esophageal cancer: A meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. Cancer Epidemiol. Biomarkers Prev. *14*, 1967–1971.

23. Wu, C., Hu, Z., He, Z., Jia, W., Wang, F., Zhou, Y., Liu, Z., Zhan, Q., Liu, Y., Yu, D., et al. (2011). Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat. Genet. *43*, 679–684.

24. García-Closas, M., Malats, N., Silverman, D., Dosemeci, M., Kogevinas, M., Hein, D.W., Tardón, A., Serra, C., Carrato, A., García-Closas, R., et al. (2005). NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: Results from the Spanish Bladder Cancer Study and meta-analyses. Lancet *366*, 649–659.

25. Kilpeläinen, T.O., Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C.H., et al. (2011). Physical activity attenuates the influence of FTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. PLoS Med. *8*, e1001116.

26. Moore, S.C., Gunter, M.J., Daniel, C.R., Reddy, K.S., George, P.S., Yurgalevitch, S., Devasenapathy, N., Ramakrishnan, L., Chatterjee, N., Chanock, S.J., et al. (2011). Common Genetic Variants and Central Adiposity Among Asian-Indians. Obesity (Silver Spring). Published online July 28, 2011.

27. Kallberg, H., Padyukov, L., Plenge, R.M., Ronnelid, J., Gregersen, P.K., van der Helm-van Mil, A.H., Toes, R.E., Huizinga, T.W., Klareskog, L., and Alfredsson, L.; Epidemiological Investigation of Rheumatoid Arthritis study group. (2007). Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am. J. Hum. Genet. *80*, 867–875.

28. Pencina, M.J., D'Agostino, R.B., Sr., and Steyerberg, E.W. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat. Med. *30*, 11–21.

29. Pencina, M.J., D'Agostino, R.B., Sr., D'Agostino, R.B., Jr., and Vasan, R.S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat. Med. *27*, 157–172, discussion 207–212.

30. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., and Mulvihill, J.J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J. Natl. Cancer Inst. *81*, 1879–1886.

31. Pepe, M.S. (2003). The Statistical Evaluation of Medical Tests for Classification and Prediction (New York: Oxford University Press).

32. Cook, N.R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation *115*, 928–935.

33. Gail, M.H., and Pfeiffer, R.M. (2005). On criteria for evaluating models of absolute risk. Biostatistics *6*, 227–239.

34. Pepe, M.S., and Janes, H.E. (2008). Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer. J. Natl. Cancer Inst. *100*, 978–979.

35. Greenland, S. (2008). The need for reorientation toward cost-effective prediction: comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by M. J. Pencina et al., Statistics in Medicine (DOI: 10.1002/sim.2929). Stat. Med. *27*, 199–206.

36. Milne, R.L., Gaudet, M.M., Spurdle, A.B., Fasching, P.A., Couch, F.J., Benítez, J., Arias Pérez, J.I., Zamora, M.P., Malats, N., Dos Santos Silva, I., et al; GENICA Network; kConFab; AOCS. (2010). Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. Breast Cancer Res. *12*, R110.

37. Qi, L., and Liang, J. (2010). Interactions between genetic factors that predict diabetes and dietary factors that ultimately impact on risk of diabetes. Curr. Opin. Lipidol. *21*, 31–37.

38. Carlborg, O., and Haley, C.S. (2004). Epistasis: Too often neglected in complex trait studies? Nat. Rev. Genet. *5*, 618–625.

39. Shao, H., Burrage, L.C., Sinasac, D.S., Hill, A.E., Ernest, S.R., O'Brien, W., Courtland, H.W., Jepsen, K.J., Kirby, A., Kulbokas, E.J., et al. (2008). Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. Proc. Natl. Acad. Sci. USA *105*, 19910–19914.

40. Bouzigon, E., Corda, E., Aschard, H., Dizier, M.H., Boland, A., Bousquet, J., Chateigner, N., Gormand, F., Just, J., Le Moual, N., et al. (2008). Effect of 17q21 variants and smoking exposure in early-onset asthma. N. Engl. J. Med. *359*, 1985–1994.

41. Mechanic, L.E., Chen, H.S., Amos, C.I., Chatterjee, N., Cox, N.J., Divi, R.L., Fan, R., Harris, E.L., Jacobs, K., Kraft, P., et al. (2011). Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. Genet. Epidemiol.

42. Bookman, E.B., McAllister, K., Gillanders, E., Wanke, K., Balshaw, D., Rutter, J., Reedy, J., Shaughnessy, D., Agurs-Collins, T., Paltoo, D., et al; for the NIH G × E Interplay Workshop participants. (2011). Gene-environment interplay in common complex diseases: Forging an integrative model-recommendations from an NIH workshop. Genet. Epidemiol.

43. Hughes, D.A., Vilar, F.J., Ward, C.C., Alfirevic, A., Park, B.K., and Pirmohamed, M. (2004). Cost-effectiveness analysis of HLA B*5701 genotyping in preventing abacavir hypersensitivity. Pharmacogenetics *14*, 335–342.

44. Eichelbaum, M., Ingelman-Sundberg, M., and Evans, W.E. (2006). Pharmacogenomics and individualized drug therapy. Annu. Rev. Med. *57*, 119–137.

45. Rose, G. (1992). The strategy of preventive medicine (Oxford: Oxford University Press).

46. So, H.C., and Sham, P.C. (2010). A unifying framework for evaluating the predictive power of genetic variants based on the level of heritability explained. PLoS Genet. *6*, e1001230.

47. Mihaescu, R., van Zitteren, M., van Hoek, M., Sijbrands, E.J., Uitterlinden, A.G., Witteman, J.C., Hofman, A., Hunink, M.G., van Duijn, C.M., and Janssens, A.C. (2010). Improvement of risk prediction by genomic profiling: reclassification measures versus the area under the receiver operating characteristic curve. Am. J. Epidemiol. *172*, 353–361.

48. Cordell, H.J. (2009). Detecting gene-gene interactions that underlie human diseases. Nat. Rev. Genet. *10*, 392–404.

49. Siemiatycki, J., and Thomas, D.C. (1981). Biological models and statistical interactions: An example from multistage carcinogenesis. Int. J. Epidemiol. *10*, 383–387.

50. Kraft, P., Yen, Y.C., Stram, D.O., Morrison, J., and Gauderman, W.J. (2007). Exploiting gene-environment interaction to detect genetic associations. Hum. Hered. *63*, 111–119.

51. Murcray, C.E., Lewinger, J.P., and Gauderman, W.J. (2009). Gene-environment interaction in genome-wide association studies. Am. J. Epidemiol. *169*, 219–226.

52. Thomas, D.C., Lewinger, J.P., Murcray, C.E., and Gauderman, W.J. (2012). Invited commentary: GE-Whiz! Ratcheting gene-environment studies up to the whole genome and the whole exposome. Am. J. Epidemiol. *175*, 203–207, discussion 208–209.

53. Mukherjee, B., Ahn, J., Gruber, S.B., and Chatterjee, N. (2012). Testing gene-environment interaction in large-scale case-control association studies: Possible choices and comparisons. Am. J. Epidemiol. *175*, 177–190.

54. Cornelis, M.C., Tchetgen, E.J., Liang, L., Qi, L., Chatterjee, N., Hu, F.B., and Kraft, P. (2012). Gene-environment interactions in genome-wide association studies: A comparative study of tests applied to empirical studies of type 2 diabetes. Am. J. Epidemiol. *175*, 191–202.

55. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.

56. International HapMap Consortium. (2003). The International HapMap Project. Nature *426*, 789–796.

57. Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. *39*, 870–874.

58. Qi, L., Cornelis, M.C., Kraft, P., Stanya, K.J., Linda Kao, W.H., Pankow, J.S., Dupuis, J., Florez, J.C., Fox, C.S., Paré, G., et al; Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC); Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. (2010). Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. Hum. Mol. Genet. *19*, 2706–2715.

59. Fernando, M.M., Stevens, C.R., Walsh, E.C., De Jager, P.L., Goyette, P., Plenge, R.M., Vyse, T.J., and Rioux, J.D. (2008). Defining the role of the MHC in autoimmunity: A review and pooled analysis. PLoS Genet. *4*, e1000024.

60. Karlson, E.W., Chibnik, L.B., Kraft, P., Cui, J., Keenan, B.T., Ding, B., Raychaudhuri, S., Klareskog, L., Alfredsson, L., and Plenge, R.M. (2010). Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Ann. Rheum. Dis. *69*, 1077–1085.

61. Karlson, E.W., Chibnik, L.B., Tworoger, S.S., Lee, I.M., Buring, J.E., Shadick, N.A., Manson, J.E., and Costenbader, K.H. (2009). Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. Arthritis Rheum. *60*, 641–652.

62. Karlson, E.W., Mandl, L.A., Hankinson, S.E., and Grodstein, F. (2004). Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. Arthritis Rheum. *50*, 3458–3467.

63. Costenbader, K.H., Feskanich, D., Mandl, L.A., and Karlson, E.W. (2006). Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. Am. J. Med. *119*, 503.e1–503.e9.

64. Willett, W.C., Stampfer, M.J., Colditz, G.A., Rosner, B.A., Hennekens, C.H., and Speizer, F.E. (1987). Moderate alcohol consumption and the risk of breast cancer. N. Engl. J. Med. *316*, 1174–1180.

65. Thomas, G., Jacobs, K.B., Kraft, P., Yeager, M., Wacholder, S., Cox, D.G., Hankinson, S.E., Hutchinson, A., Wang, Z., Yu, K., et al. (2009). A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat. Genet. *41*, 579–584.

66. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.