BMC
Cardiovascular Disorders

RESEARCH ARTICLE   **Open Access**

Genotype-informed estimation of risk of coronary heart disease based on genome-wide association data linked to the electronic medical record

Keyue Ding[1], Kent R Bailey[2] and Iftikhar J Kullo[1*]

Abstract

Background: Susceptibility variants identified by genome-wide association studies (GWAS) have modest effect sizes. Whether such variants provide incremental information in assessing risk for common 'complex' diseases is unclear. We investigated whether measured and imputed genotypes from a GWAS dataset linked to the electronic medical record alter estimates of coronary heart disease (CHD) risk.

Methods: Study participants ($n = 1243$) had no known cardiovascular disease and were considered to be at high, intermediate, or low 10-year risk of CHD based on the Framingham risk score (FRS) which includes age, sex, total and HDL cholesterol, blood pressure, diabetes, and smoking status. Of twelve SNPs identified in prior GWAS to be associated with CHD, four were genotyped in the participants as part of a GWAS. Genotypes for seven SNPs were imputed from HapMap CEU population using the program MACH. We calculated a multiplex genetic risk score for each patient based on the odds ratios of the susceptibility SNPs and incorporated this into the FRS.

Results: The mean (SD) number of risk alleles was 12.31 (1.95), range 6-18. The mean (SD) of the weighted genetic risk score was 12.64 (2.05), range 5.75-18.20. The CHD genetic risk score was not correlated with the FRS ($P = 0.78$). After incorporating the genetic risk score into the FRS, a total of 380 individuals (30.6%) were reclassified into higher-(188) or lower-risk groups (192).

Conclusion: A genetic risk score based on measured/imputed genotypes at 11 susceptibility SNPs, led to significant reclassification in the 10-y CHD risk categories. Additional prospective studies are needed to assess accuracy and clinical utility of such reclassification.

Background

Genome-wide association studies (GWAS) have identified multiple SNPs as being associated with the risk of developing common 'complex' diseases [1-9]. The potential use of susceptibility SNPs in individual-level risk estimation and clinical decision-making is a topic of considerable interest [10-12]. Since effect sizes (ie, odds ratio or relative risk) of susceptibility SNPs for common diseases are modest, whether such variants provide incremental risk prediction beyond conventional risk prediction algorithms is unclear. Of note, direct-to-consumer (DTC) companies are already providing genotype based estimates of risk of common diseases in the absence of established clinical utility [13].

The 10-year (10-y) risk of coronary heart disease (CHD) is estimated based on conventional risk factors using the Framingham risk score (FRS) [14], and enables preventive measures to be targeted toward individuals who need these the most. Such individuals can be treated by lifestyle modification and/or drug therapy [15]. The FRS is based on age, sex, diabetes, smoking, blood pressure categories, and total and high-density lipoprotein (HDL) cholesterol levels.

A SNP on chromosome 9p21 has been tested in several studies for its utility in refining estimates of CHD risk, but so far the results are inconsistent [16-18]. Nine SNPs that influence serum lipid levels were associated with adverse cardiovascular events but did not improve discrimination and only slightly improved reclassification [19].

* Correspondence: kullo.iftikhar@mayo.edu
[1]Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA
Full list of author information is available at the end of the article

BioMed Central

A genetic risk score based on 101 SNPs associated with cardiovascular disease phenotypes and related intermediate phenotypes was not significantly associated with incident adverse cardiovascular events after adjustment for conventional cardiovascular risk factors [20]. However, in a propective cohort and case-control analysis, a genetic risk score based on 13 SNPs associated with CHD identified 20% of the participants who were at ~70% increased risk of a first CHD event [21].

Electronic medical record (EMR)-based GWAS have been proposed to overcome the bottleneck of high-phenotyping costs and thereby facilitate genomic studies of diverse medically relevant phenotypes. We investigated whether measured and imputed genotypes from a GWAS dataset linked to the EMR alter estimates of CHD in 1243 individuals from the Mayo electronic Medical Records and Genomics (eMERGE) cohort [22,23], which comprises of peripheral arterial disease cases and controls. We calculated a multiplex genetic risk score, incorporated it into the FRS, and then assessed extent of subsequent reclassification of CHD risk.

Methods

Study participants
In the Mayo eMERGE cohort, peripheral arterial disease cases had an ankle-brachial index (ABI) of ≤0.9 at rest or one minute after exercise; or the presence of poorly compressible arteries; or a history of lower extremity revascularization. Controls were identified from patients referred to the Cardiovascular Health Clinic for CHD screening and had no history of peripheral arterial disease. For the present study, we excluded control patients who had CHD, defined as the presence of the International Classification of Disease-9-Clinical Modification (ICD-9-CM) diagnosis codes 410.33-414.33, or a history of percutaneous coronary intervention or coronary artery bypass surgery (ICD-9-CM procedure codes 36.10-36.14). In all, 1243 controls without known cardiovascular disease were identified.

All participants gave their written informed consent for participation in the study and the use of their data for future research. The study protocol was approved by the Institutional Review Board of the Mayo Clinic.

Genetic marker selection and imputation
At the time of conducting this study, 12 SNPs were reported to be associated with CHD (myocardial infarction and sudden cardiac death) in GWAS [24]. The 12 susceptibility genes, the corresponding SNPs, risk allele, risk allele frequencies (RAFs), and the sizes of their effects (ie, odds ratio and 95% confidence interval) are listed in Table 1. We used fixed-effects models to calculate the summary odds ratio and 95% confidence intervals for four out of 12 SNPs (rs599839, rs501120, rs1746048, and rs3008621) based on published studies (reviewed in Kullo and Cooper [24]). For the remaining SNPs, the odds ratios and 95% confidence intervals were derived from the combined or pooled analyses in the original studies.

Of the 12 SNPs, four (rs3184504, rs6725887, rs11206510, and rs1746048) were genotyped on the Human660W-Quad v1 chip used in the Mayo eMERGE study. The quality control procedures adopted in the eMERGE network have been described elsewhere [25]. The following criteria were used: SNP call rate > 98%, sample call rate > 98%, minor allele frequency > 0.05, Hardy-Weinberg equilibrium $P > 0.001$, and 99.99% concordance rate in duplicates. Of the remaining eight SNPs, genotypes for seven were imputed; however the genotypes for SNP rs3008621 could not be imputed. Thus, 11 out of the 12 susceptibility SNPs were used in the subsequent analyses. We used the program MACH [26,27], and haplotypes derived from the HapMap II CEU samples for imputation. The quality of imputation

Table 1 SNPs associated with CHD in genome-wide association studies

SNP	Chr.	Gene	Risk allele	ORs (95% CI)	Risk allele frequency	
					Reported	Observed
rs10757278	9	*CDKN2A/2B*	G	1.28 (1.22, 1.35)	0.45	0.52
rs599839	1	*SORT1*	A	1.17 (1.11, 1.22)	0.78	0.78
rs3008621	1	*MIA3*	G	1.11 (1.04, 1.19)	0.26	NA
rs501120	10	*Unknown*	T	1.30 (1.12, 1.51)	0.84	0.87
rs9818870	3	*MRAS*	T	1.15 (1.11, 1.19)	0.15	0.15
rs3184504	12	*SH2B3*	T	1.13 (1.08, 1.18)	0.39	0.49
rs9982601	21	*MRPS6*	T	1.20 (1.14, 1.27)	0.13	0.14
rs12526453	6	*PHACTR1*	C	1.12 (1.08, 1.17)	0.65	0.67
rs6725887	2	*WDR12*	C	1.17 (1.11, 1.23)	0.14	0.13
rs1122608	19	*LDLR*	G	1.15 (1.10, 1.21)	0.75	0.77
rs11206510	1	*PCSK9*	T	1.15 (1.10, 1.21)	0.81	0.82
rs1746048	10	*CXCL12*	C	1.16 (1.09, 1.24)	0.84	0.87

was assessed by the average posterior probability for the most likely genotype and the correlation coefficient R^2. The minimum average posterior probability for imputed genotypes of seven SNPs was 0.92 and the R^2 was 0.90.

Construction of the genetic risk score

To generate a genetic risk score for each individual, we assumed an additive genetic model in which the genotypes are coded '0' for non-risk allele homozygotes, '1' for heterozygotes, and '2' for risk-allele homozygotes. The genetic risk score based on the number of risk alleles was calculated as:

$$GRS = \sum n_i, \tag{1}$$

where n_i is the number of risk alleles for SNP i. A rescaled weighted genetic risk score (r_GRS_W) was calculated by multiplying the logarithm of odds ratio (w_i) by 0, 1, or 2 according to the number of risk alleles carried by each person and rescaled by the rescaling factor $k / \sum_i w_i$ [28] as:

$$r\_GRS\_W = \frac{k}{\sum_i w_i} \sum_i w_i \times n_i, \tag{2}$$

where k is the number of SNPs ($k = 11$).

Estimating the genotype effects of multiple SNPs

When combining multiple SNPs, we estimated the logarithm of the combined odds ratios for each individual relative to the average in the population, as follows:

$$\gamma_G = \sum_i (n_i - 2p_i) w_i. \tag{3}$$

Intuitively, γ_G sums the difference between observed (n_i) and expected ($2p_i$) risk allele counts across SNPs, weighted by the logarithm of odds ratio.

We also estimated the combined risk from multiple SNPs relative to the general population as follows [29]. Under the assumption of low probability of incident events, the average population risk for SNP i (R_i) relative to homozygosity for the non-risk allele, can be approximated in the multiplicative model as,

$$R_i = e^{2w_i} p_i^2 + 2e^{w_i} p_i (1 - p_i) + (1 - p_i)^2. \tag{4}$$

For each SNP, we expressed the population average risk [$pRR(n_i)$] relative to the risk for a person with zero copies of the risk allele as:

$$pRR(n_i) = \begin{cases} e^{2w_i}/R_i, & n_i = 2; \\ e^{w_i}/R_i, & n_i = 1; \\ 1/R_i, & n_i = 0. \end{cases} \tag{5}$$

where n_i is the number of risk alleles for each genotype. Therefore, the combined relative risk from multiple SNPs can be approximated assuming independent effects between SNPs as,

$$pRR = \prod_i pRR(n_i). \tag{6}$$

Incorporating the genetic risk score into the FRS

We used the method of Wilson et al. [14] to calculate the FRS and the 10-y risk of CHD (base model). Conventional risk factors for cardiovascular disease, including age, sex, total, LDL, and HDL cholesterol, blood pressure, diabetes, and smoking status (Table 2), were extracted from the Mayo EMR as previously described [22]. The electronic phenotyping algorithms had an accuracy of > 90% [22], using manual medical record review as the gold standard. The 10-y CHD risk was defined as:

$$P = 1 - s(10)^{e^A}, \tag{7}$$

where e^A represents the Framingham-based hazard ratio for CHD and $s(t)$ is the survival function. The relative hazard for CHD (e^A) was calculated from a Cox-regression model for conventional risk factors (ie, age, sex, total and HDL cholesterol, blood pressure, diabetes, and smoking status) [14].

In order to incorporate genotypes at risk alleles into the estimation of 10-y CHD risk (base model plus genetic score), we added the combined effects from multiple SNPs into the survival function as:

$$P' = 1 - s(10)^{e^{A+G}}, \tag{8}$$

where G is the combined effect from the multiple SNPs: ie, γ_G from Eq. (3) or $log_e pRR$ where pRR is from Eq. (6). In this calculation, we considered γ_G & $log_e pRR$ each as approximations of the genetic hazard ratios.

Table 2 Sample characteristics

Characteristics	Mean/ Count	SD/Proportion (%)
Sex (female)	503	39.9
Age (y)	60.2	7.1
Total cholesterol (mg/dL)	205.6	36.8
High-density lipoprotein cholesterol (mg/dL)	57.0	17.5
Low-density lipoprotein cholesterol (mg/dL)	121.0	32.4
Systolic blood pressure (mmHg)	128.1	17.5
Diastolic blood pressure (mmHg)	77.7	15.2
Diabetes	144	12.9
Smoker	468	37.1

Reclassification using genetic risk scores

Reclassification refers to the proportion of persons who change risk categories when prediction models are updated to incorporate new biomarkers [13]. We measured how often individuals were estimated to be in different risk categories when the genetic risk score was incorporated into Framingham risk score. The risk categories were defined as less than 5% risk (low), 5% to less than 10% risk (intermediate), 10% to less than 20% risk (intermediate high), and 20% or higher risk (high). This was done for both versions of the genetic risk score (γ_G and log_epRR). In addition, we repeated these calculations substituting lower and upper confidence limits for all of the genetic effects. All analyses were performed using *R* (http://www.r-project.org).

Results

Genetic risk scores

We included 11 replicated SNPs associated with CHD at a genome-wide significance level of 5×10^{-8} [24]. The reported effect sizes (ie, odds ratios), and RAFs are shown in Table 1. The RAFs in the published reports were similar to the RAFs in our sample. The mean (SD) number of risk alleles was similar: 12.31 (1.95) with a range of 6 to 18. The mean (SD) of the rescaled weighted genetic risk score was 12.64 (2.05) with a range from 5.75 to 18.20. The number of risk alleles and the weighted genetic risk scores were normally distributed (Figure 1a &1b) and there was a direct correlation ($r = 0.96$) between the two (Figure 1c). However, variability was noted in the weighted genetic risk score for each category of cumulative risk alleles, reflecting different odds ratios associated with the susceptibility SNPs. We tested the association of three SNPs known to be related to lipid traits - rs11206510 in *PCSK9*, rs599839 in *SORT1*, and rs1122608 in *LDLR* - with total cholesterol, LDL cholesterol, and HDL cholesterol;

rs11206510 was associated with total cholesterol ($P = 0.026$) and LDL cholesterol ($P = 0.027$), and rs599839 was marginally associated with HDL cholesterol ($P = 0.050$) and LDL cholesterol ($P = 0.078$). The genetic risk scores were not correlated with other conventional risk factors - systolic and diastolic blood pressure, diabetes, as well as the FRS ($P > 0.05$ for each).

Effects of combining risk SNPs

The odds ratios and risk relative to the general population increased with the number of risk alleles (Figure 2). As shown in Figure 2, the 25th and 75th percentile of the combined odds ratio (0.77 and 1.26), corresponded to the presence of 11 and 14 risk alleles, respectively. If the number of risk alleles was ≤ 11, the combined odds ratio or risk relative to the general population was < 1, indicating that individuals with ≤ 11 risk alleles had lesser risk relative to the average risk in the population. Conversely, if the number of risk alleles was ≥ 14, the combined odds ratio or risk relative to the general population was > 1. The risk relative to the general population (*pRR*) estimated from the 11 SNPs (Eq. 6) was highly correlated ($r = 0.99$) with the combined odds ratio (e^{γ_G}) obtained from Eq. (3).

Estimation of 10-y risk of CHD

The 10-y CHD risk was estimated based on the FRS and then revised after inclusion of genotype effects from susceptibility SNPs. The two genetic risk scores were not correlated with FRS ($P = 0.824$ and $P = 0.779$, respectively). A positive correlation ($r = 0.36$) was noted between the weighted genetic risk scores and the 10-y CHD risk after the inclusion of genotypes.

Reclassification using γ_G

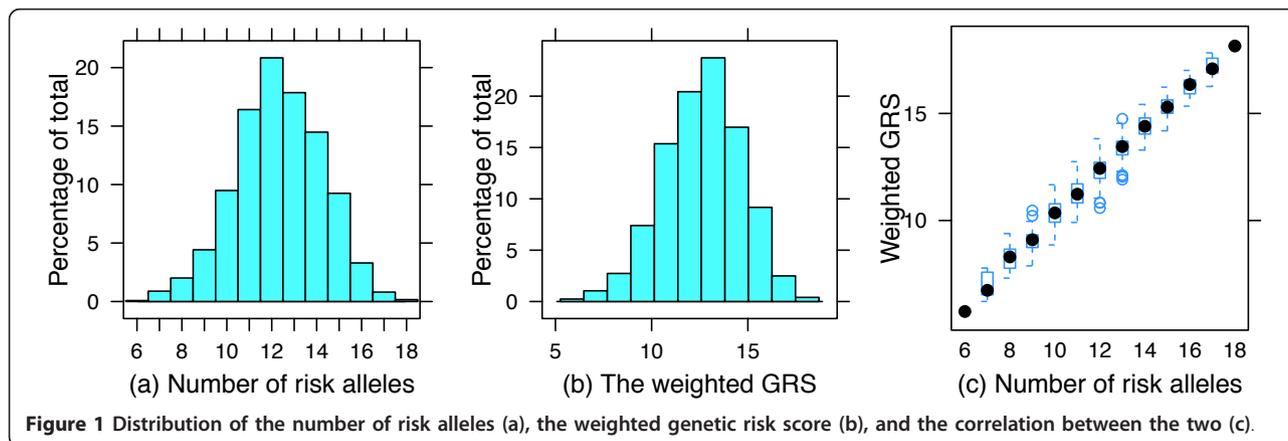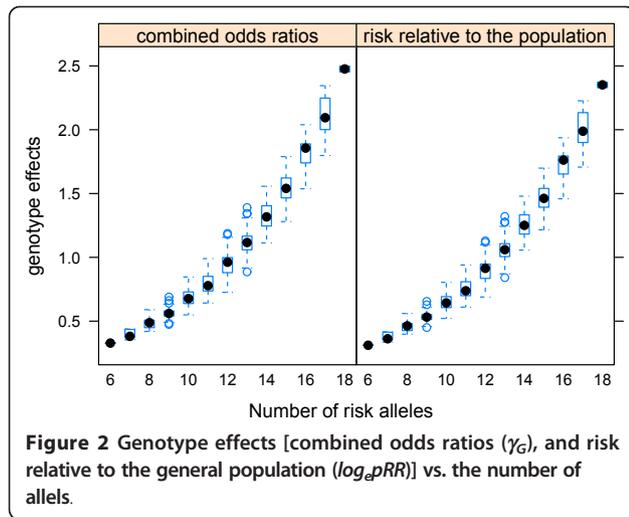We next examined reclassification of 10-y CHD risk after incorporating γ_G from Eq. 3 (Table 3). Nearly all of

Figure 1 Distribution of the number of risk alleles (a), the weighted genetic risk score (b), and the correlation between the two (c).

Figure 2 Genotype effects [combined odds ratios (γ_G), and risk relative to the general population ($log_e pRR$)] vs. the number of allels.

the individuals reclassified were placed into adjacent risk categories. Only one patient in the intermediate-high group was reclassified into the low-risk group. In all, 188 and 192 individuals were reclassified into higher-and lower-risk groups, respectively, a reclassification rate of 30.6%. The reclassification rate was higher in those whose FRS placed them in the intermediate category: 41.3% (162 out of 392) were reclassified. The reclassification rate in the intermediate-high category was 30.8% (130 out of 422), similar to the overall reclassification rate. The reclassification rate in the low- (25.5%; 50 out of 196) and high- (16.3%; 38 out of 233) categories was lower than the overall reclassification rate. We also estimated the uncertainty of reclassification by repeating the

above calculations using the lower and upper 95% confidence levels of odds ratio for each genetic variant (Table 3). The reclassification rate was 22.2% (130 and 146 were reclassified into high- and low-risk groups) and 38.6% (239 and 241 were reclassified into high- and low-risk groups) for the lower and upper limits of the odds ratios, respectively.

Reclassification using $log_e pRR$

The reclassification rate of 10-y CHD risk after incorporating the multiplex genetic risk score $log_e pRR$ (where pRR is from Eq. 6) is shown in Table 4. For each risk category, the pattern of the reclassification rate was similar to that using γ_G. All of the individuals reclassified were placed into adjacent risk categories except two patients in the intermediate-high group who were reclassified into the low-risk group. In all, 154 and 228 individuals were reclassified into higher- and lower- risk groups (reclassification rate 30.7%), indicating fewer individuals were reclassified into higher-risk category. Using the lower and upper 95% confidence levels of odds ratios in the calculation of pRR, the reclassification rate was 22.1% (113 and 162 were reclassified into high- and low-risk groups) and 38.7% (184 and 297 were reclassified into high- and low-risk groups), respectively.

Discussion

Genome-wide association studies (GWAS) have identified multiple susceptibility variants for common 'complex' diseases. However, the clinical utility of these variants and whether GWAS results should be communicated to study participants is not clear [30]. The

Table 3 Reclassification of 10-y CHD risk after the addition of genotype information (γ_G)

10-y CHD risk	FRS	Risk	FRS+γ_G			
			Low	INT	INT-high	High
Low (≤5%)	196 (15.8%)	ORs	146	50	0	0
		Lower	156	40	0	0
		Upper	138	58	0	0
INT (>5%,≤10%)	392 (31.5%)	ORs	76	230	86	0
		Lower	55	285	52	0
		Upper	91	200	98	3
INT-high (>10%,≤20%)	422 (34.0%)	ORs	1	77	292	52
		Lower	0	62	322	38
		Upper	5	98	239	80
High (>20%)	233 (18.7%)	ORs	0	0	38	195
		Lower	0	0	29	204
		Upper	0	0	45	186
Total	1243	ORs	223 (17.9%)	357 (28.7%)	416 (33.5%)	247 (19.9%)
		Lower	211 (17.0%)	387 (31.3%)	403 (32.4%)	242 (19.5%)
		Upper	234 (18.8%)	358 (28.8%)	382 (30.7%)	269 (21.6%)

INT: Intermediate; γ_G: combined odds ratio from Eq. (3)

Lower and Upper: the lower and upper 95% confidence level for odds ratios used to calculate γ_G

Table 4 Reclassification of 10-y CHD risk after the addition of genotype information (*log$_e$pRR*)

10-y CHD risk	FRS	Risk	FRS+*log$_e$pRR*			
			Low	INT	INT-high	High
Low (≤5%)	196 (15.8%)	pRR	151	45	0	0
		Lower	161	35	0	0
		Upper	148	48	0	0
INT (>5%,≤10%)	392 (31.5%)	pRR	88	240	64	0
		Lower	59	288	45	0
		Upper	112	199	81	0
INT-high (>10%,≤20%)	422 (34.0%)	pRR	1	90	286	45
		Lower	0	67	322	33
		Upper	11	115	241	55
High (>20%)	233 (18.7%)	pRR	0	0	49	184
		Lower	0	0	36	197
		Upper	0	6	53	174
Total	1243	pRR	240 (19.3%)	375 (30.2%)	399 (32.1%)	229 (18.4%)
		Lower	220 (17.7%)	390 (31.4%)	403 (32.4%)	230 (18.5%)
		Upper	271 (21.8%)	368 (29.6%)	375 (30.2%)	229 (18.4%)

INT: Intermediate; pRR, risk relative to the general population from Eq. (6)
Lower and Upper: the lower and upper 95% confidence level for odds ratios used to calculate *pRR*

presence of multiple genetic susceptibility variants in an individual may lead to additive, clinically relevant increases in risk of disease and such knowledge may refine risk stratification. In the present study of 1243 individuals without known CHD who had undergone high-density genome-wide genotyping, we found that incorporating genotypes at 11 susceptibility SNPs into the FRS led to significant reclassification of the estimated 10-y CHD risk. Our study demonstrates that GWAS genotypes linked to an EMR can be used to create a multiplex genetic risk score - based on both measured/imputed genotypes - that in-turn can be used to revise the risk estimated based on conventional risk factors. Most of the selected SNPs were associated with CHD but not with related risk factors, and therefore the CHD genetic risk score was not correlated with FRS.

It should be noted that individualized risk estimates are based on statistical modeling of population level data, and have a level of uncertainty due to variation in risk allele frequency, effect sizes, and modeling of combined effects [31]. Physicians and patients should be aware of this and avoid attributing an unreasonable degree of certainty to such estimates, which are often presented as a single number, ie, 'the probability of occurrence of a negative genetic outcome' [32]. Additionally, probabilities and risks may be difficult to interpret, and patients often have poor understanding and recall of objective risk estimates regardless of the format in which they were presented and conveyed. This highlights a need for developing tools to better communicate genomic components of disease risk to patients.

Direct-to-consumer (DTC) companies, such as 23andMe [33], deCODEme [34], and Navigenics [35], disclose risk estimates to 'customers' as soon as initial GWAS are published; instantaneous translation of multiple risk markers is of unclear clinical benefit and may be harmful. The three companies use different approaches to generate a genetic risk score although all three assume a multiplicative model. 23andMe takes the product of relative risks of all SNPs and multiplies this value by the average population risk to generate an estimate of the individual's lifetime risk [33]. Similar to what is shown in Eqs. 4-6, deCODEme applies risk of each SNP to the population and then takes the product. Navigenics generates an interim 'genetic composite index' number, which incorporates known risk factors, as well as other information and assumptions such as the allele frequencies and the prevalence of the disease [36].

We noted significant reclassification (30.6%) using combined odds ratios of the susceptibility SNPs. If risk categories are used to define thresholds for type or intensity of interventions (eg, cholesterol-lowering drugs), reclassification can impact clinical management. In the context of the present study, treatment consideration may change in patients reclassified as high-risk for CHD [18]. Additional studies are needed to demonstrate whether the results of genetic testing (eg, a genetic risk score) motivate patients to make lifestyle changes, whether physicians understand genetic risk and make decisions based on the risk, and whether genetic testing improves outcomes in selected patient populations. The calibration and discriminative accuracy of genetic risk

scores needs to be assessed in prospective studies. So far such studies have revealed poor discriminative capacity of genetic risk scores, but potentially useful reclassification [18,20,21,37].

The common variants identified in GWAS have modest effect sizes and explain only a small proportion of heritable risk. However, since the risk alleles are common, the population attributable risk is significant. It has been suggested that rare variants make a substantial contribution to the overall multifactorial inheritance of a disease [38]. Sequencing of exomes and whole genomes is being used to detect rare variants that mediate susceptibility to common 'complex' diseases, and we anticipate incorporation of such variants in disease risk scores in the near future.

Limitations

Several challenges arise in the use of GWAS genotypes for clinical application. First, different genotyping platforms are employed in GWAS and not all susceptibility SNPs may have been genotyped on a single platform. Genotypes for SNPs that were not genotyped directly have to be imputed. This is typically done using the HapMap database and several algorithms [26], which although not perfect, are highly accurate (R^2 = 0.90 for MACH). Second, an issue relevant to the potential clinical use of the genotypes from a GWAS is that genotyping is typically not performed in a Clinical Laboratory Improvement Amendments (CLIA) certified laboratory. Third, we used odds ratios since hazard ratios were not available from the case-control GWAS. Odds ratios over-estimate the relative risk in common 'complex' diseases. Fourth, we assumed an additive model when incorporating genotype effects into the risk prediction. However, other models (eg, multiplicative, recessive, or dominant) may be operative in common 'complex' diseases. Fifth, genetic risk scores will need to be periodically updated as new susceptibility variants are identified. Sixth, because the susceptibility SNPs were identified in adults of European ancestry, use of these SNPs in other populations would be problematic because the associations with CHD may not be present in other ethnic groups. The approach described here is applicable to patients of diverse ethnicities once susceptibility SNPs for these ethnic groups are identified. Seventh, incorporating family history in the EMR and integrating it with multiplex genetic risk scores needs additional work. The Center for Disease Control and Prevention has created a 'Family Healthware' software, which is a family history-screening tool for common chronic diseases and that can be incorporated into the EMR [39]. Finally, additional prospective studies are needed to confirm whether susceptibility SNPs identified in GWAS improve the accuracy of CHD risk stratification and whether multiplex genetic testing has clinical utility.

Conclusions

This study demonstrates the use of genotypes from an EMR-based GWAS to construct a multiplex genetic risk score and revise the estimated risk of a common disease - CHD. A genetic risk score based on genotypes at 11 susceptibility SNPs led to significant reclassification in the 10-y CHD risk. However, the cross sectional nature of the present study does not allow us to quantify the accuracy of risk reclassification. Additional prospective studies are needed to confirm whether susceptibility SNPs identified in GWAS improve the accuracy of CHD risk stratification and whether multiplex genetic risk scores for common diseases have clinical utility.

Author details
[1]Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA. [2]Division of Biomedical Informatics and Statistics, Mayo Clinic, Rochester, MN 55905, USA.

Authors' contributions
Conception and design: IJK. Data analyses: KD and KRB. Manuscript preparation: KD, KRB and IJK. All authors have given final approval of the manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson D, Magnusson K, Andersen K, Levey A, Backman V, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper W, Reilly M, Granger C, Austin H, Rader D, Shah S, Quyyumi A, Gulcher J, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K: **A common variant on chromosome 9p21 affects the risk of myocardial infarction.** *Science* 2007, **316(5830)**:1491-1493.
2. The Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447(7145)**:661-678.
3. Samani N, Erdmann J, Hall A, Hengstenberg C, Mangino M, Mayer B, Dixon R, Meitinger T, Braund P, Wichmann H, Barrett J, Konig I, Stevens S, Szymczak S, Tregouet D, Iles M, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth A, Baessler A, Ball S, Strom T, Braenne I, Gieger C, Deloukas P, Tobin M, Ziegler A, Thompson J, Schunkert H: **Genomewide association analysis of coronary artery disease.** *N Engl J Med* 2007, **357(5)**:443-453.
4. Coronary Artery Disease Consortium, Samani NJ, Deloukas P, Erdmann J, Hengstenberg C, Kuulasmaa K, McGinnis R, Schunkert H, Soranzo N, Thompson J, Tiret L, Ziegler A: **Large scale association analysis of novel genetic loci for coronary artery disease.** *Arterioscler Thromb Vasc Biol* 2009, **29(5)**:774-780.
5. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU,

Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PIW, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA: **Common variants at 30 loci contribute to polygenic dyslipidemia.** Nat Genet 2009, **41**:56-65.

6. Gudbjartsson DF, Bjornsdottir US, Halapi E, Helgadottir A, Sulem P, Jonsdottir GM, Thorleifsson G, Helgadottir H, Steinthorsdottir V, Stefansson H, Williams C, Hui J, Beilby J, Warrington NM, James A, Palmer LJ, Koppelman GH, Heinzmann A, Krueger M, Boezen HM, Wheatley A, Altmuller J, Shin HD, Uh ST, Cheong HS, Jonsdottir B, Gislason D, Park CS, Rasmussen LM, Porsbjerg C, Hansen JW, Backer V, Werge T, Janson C, Jönsson UB, Ng MCY, Chan J, So WY, Ma R, Shah SH, Granger CB, Quyyumi AA, Levey AI, Vaccarino V, Reilly MP, Rader DJ, Williams MJA, van Rij AM, Jones GT, Trabetti E, Malerba G, Pignatti PF, Boner A, Pescollderungg L, Girelli D, Olivieri O, Martinelli N, Ludviksson BR, Ludviksdottir D, Eyjolfsson GI, Arnar D, Thorgeirsson G, Deichmann K, Thompson PJ, Wjst M, Hall IP, Postma DS, Gislason T, Gulcher J, Kong A, Jonsdottir I, Thorsteinsdottir U, Stefansson K: **Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction.** Nat Genet 2009, **41(3)**:342-347.

7. Willer C, Sanna S, Jackson A, Scuteri A, Bonnycastle L, Clarke R, Heath S, Timpson N, Najjar S, Stringham H, Strait J, Duren W, Maschio A, Busonero F, Mulas A, Albai G, Swift A, Morken M, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen W, Li Y, Scott L, Scheet P, Sundvall J, Watanabe R, Nagaraja R, Ebrahim S, Lawlor D, Ben-Shlomo Y, Davey-Smith G, Shuldiner A, Collins R, Bergman R, Uda M, Tuomilehto J, Cao A, Collins F, Lakatta E, Lathrop G, Boehnke M, Schlessinger D, Mohlke K, Abecasis G: **Newly identified loci that influence lipid concentrations and risk of coronary artery disease.** Nat Genet 2008, **40(2)**:161-169.

8. Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Trégouët DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, Mokhtari NEE, Schäfer A, März W, Renner W, Bugert P, Klüter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Italian Atherosclerosis, Thrombosis, and Vascular Biology Working Group, Myocardial Infarction Genetics Consortium and Wellcome Trust Case Control Consortium, Consortium Cardiogenics, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H: **New susceptibility locus for coronary artery disease on chromosome 3q22.3.** Nat Genet 2009, **41(3)**:280-282.

9. Ding K, Kullo IJ: **Genome-wide association studies for atherosclerotic vascular disease and its risk factors.** Circ Cardiovasc Genet 2009, **2**:63-72.

10. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE: **Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers.** PLoS Genet 2009, **5(2)**:e1000337.

11. Damani SB, Topol EJ: **Future use of genomics in coronary artery disease.** J Am Coll Cardiol 2007, **50(20)**:1933-1940.

12. Wray N, Goddard M, Visscher P: **Prediction of individual genetic risk of complex disease.** Curr Opin Genet Dev 2008, **18(3)**:257-263.

13. Khoury MJ, McBride CM, Schully SD, Ioannidis JPA, Feero WG, Janssens ACJW, Gwinn M, Simons-Morton DG, Bernhardt JM, Cargill M, Chanock SJ, Church GM, Coates RJ, Collins FS, Croyle RT, Davis BR, Downing GJ, Duross A, Friedman S, Gail MH, Ginsburg GS, Green RC, Greene MH, Greenland P, Gulcher JR, Hsu A, Hudson KL, Kardia SLR, Kimmel PL, Lauer MS, Miller AM, Offit K, Ransohoff DF, Roberts JS, Rasooly RS, Stefansson K, Terry SF, Teutsch SM, Trepanier A, Wanke KL, Witte JS, Xu J: **The Scientific Foundation for personal genomics: recommendations from a National Institutes of Health-Centers for Disease Control and Prevention multidisciplinary workshop.** Genet Med 2009, **11(8)**:559-567.

14. Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W: **Prediction of coronary heart disease using risk factor categories.** Circulation 1998, **97(18)**:1837-47.

15. Evaluation Expert Panel on Detection and Treatment of High Blood Cholesterol in Adults: **Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III).** JAMA 2001, **285(19)**:2486-2497.

16. Talmud P, Cooper J, Palmen J, Lovering R, Drenos F, Hingorani A, Humphries S: **Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men.** Clin Chem 2008, **54(3)**:467-474.

17. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM: **Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3.** Ann Intern Med 2009, **150(2)**:65-72.

18. Brautbar A, Ballantyne CM, Lawson K, Nambi V, Chambless L, Folsom AR, Willerson JT, Boerwinkle E: **Impact of adding a single allele in the 9p21 locus to traditional risk factors on reclassification of coronary heart disease risk and implications for lipid-modifying therapy in the Atherosclerosis Risk in Communities study.** Circ Cardiovasc Genet 2009, **2(3)**:279-285.

19. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt N, Roos C, Hirschhorn J, Berglund G, Hedblad B, Groop L, Altshuler D, Newton-Cheh C, Orho-Melander M: **Polymorphisms associated with cholesterol and risk of cardiovascular events.** N Engl J Med 2008, **358(12)**:1240-1249.

20. Paynter NP, Chasman DI, Paré G, Buring JE, Cook NR, Miletich JP, Ridker PM: **Association between a literature-based genetic risk score and cardiovascular events in women.** JAMA 2010, **303(7)**:631-637.

21. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki ML, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S: **A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses.** Lancet 2010, **376(9750)**:1393-1400.

22. Kullo IJ, Fan J, Pathak J, Savova GK, Ali Z, Chute CG: **Leveraging informatics for genetic studies: use of the electronic medical record to enable a genome-wide association study of peripheral arterial disease.** J Am Med Inform Assoc 2010, **17(5)**:568-574.

23. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG: **A genome-wide association study of red blood cell traits using the electronic medical record.** PLoS ONE 2010, **5(9)**:e13011.

24. Kullo IJ, Cooper LT: **Early identification of cardiovascular risk using genomics and proteomics.** Nat Rev Cardiol 2010, **7(6)**:309-317.

25. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, Andrade MD, Doheny KF, nathan L, Haines J, Hayes G, Jarvik G, Jiang L, Kullo IJ, Li R, Ling H, Manolio TA, Matsumoto M, McCarty CA, McDavid AN, Mirel DB, Paschall JE, Pugh EW, Rasmussen LV, Wilke RA, Zuvich RL, Ritchie MD: **Quality control procedures for genome-wide association studies.** Curr Protoc Hum Genet 2011, , **68**: 1.19-19.18, Chapter 1:Unit1.19.

26. Li Y, Willer C, Sanna S, Abecasis G: **Genotype imputation.** Annual review of genomics and human genetics 2009, **10**:387-406.

27. MACH program. [http://www.sph.umich.edu/csg/abecasis/mach/index.html].

28. Lin X, Song K, Lim N, Yuan X, Johnson T, Abderrahmani A, Vollenweider P, Stirnadel H, Sundseth SS, Lai E, Burns DK, Middleton LT, Roses AD, Matthews PM, Waeber G, Cardon L, Waterworth DM, Mooser V: **Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score-the CoLaus Study.** Diabetologia 2009, **52(4)**:600-608.

29. Risk Calculation in deCODEme. [http://www.decodeme.com/health-watch-information/risk-calculation].

30. Thanassoulis G, Vasan RS: **Genetic cardiovascular risk prediction: will we get there?** Circulation 2010, **122(22)**:2323-2334.

31. Yang Q, Flanders WD, Moonesinghe R, Ioannidis JPA, Guessous I, Khoury MJ: **Using lifetime risk estimates in personal genomic profiles: estimation of uncertainty.** Am J Hum Genet 2009, **85(6)**:786-800.

32. Shiloh S: **Decision-making in the context of genetic risk.** In The troubled helix: Social and psychological implications of the new human genetics.. xvii edition. Edited by: Marteau T, Martin R. New York, NY: Cambridge: Cambridge University; 1996:82-103.

33. The 23andMe website. [http://www.23andme.com].

34. The deCODEme website. [http://www.decodeme.com].

35. The Navigenics website. [http://www.navigenics.com].
36. Genetic composite index in Navigenics. [http://www.navigenics.com/
    static/pdf/Navigenics-TheScience.pdf].
37. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR,
    Thun MJ, Cox DG, Hankinson SE, Kraft P, Rosner B, Berg CD, Brinton LA,
    Lissowska J, Sherman ME, Chlebowski R, Kooperberg C, Jackson RD,
    Buckman DW, Hui P, Pfeiffer R, Jacobs KB, Thomas GD, Hoover RN, Gail MH,
    Chanock SJ, Hunter DJ: Performance of common genetic variants in
    breast-cancer risk models. *N Engl J Med* 2010, **362(11)**:986-993.
38. Bodmer W, Bonilla C: Common and rare variants in multifactorial
    susceptibility to common diseases. *Nat Genet* 2008, **40(6)**:695-701.
39. Family Healthware. [http://www.cdc.gov/genomics/famhistory/famhx.htm].

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2261/11/66/prepub