

Improved prediction of complex diseases by common genetic markers: state of the art and further perspectives

Bent Müller¹ · Arndt Wilcke^{1,2} · Anne-Laure Boulesteix³ · Jens Brauer⁴ · Eberhard Passarge^{5,6} · Johannes Boltze^{1,2,7,8} · Holger Kirsten^{1,2,9}

Received: 16 September 2015 / Accepted: 15 January 2016 / Published online: 2 February 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Reliable risk assessment of frequent, but treatable diseases and disorders has considerable clinical and socio-economic relevance. However, as these conditions usually originate from a complex interplay between genetic and environmental factors, precise prediction remains a considerable challenge. The current progress in genotyping technology has resulted in a substantial increase of knowledge regarding the genetic basis of such diseases and disorders. Consequently, common genetic risk variants are increasingly being included in epidemiological models to improve risk prediction. This work reviews recent high-quality publications targeting the prediction of common complex diseases. To be included in this review, articles had to report both, numerical measures of prediction

performance based on traditional (non-genetic) risk factors, as well as measures of prediction performance when adding common genetic variants to the model. Systematic PubMed-based search finally identified 55 eligible studies. These studies were compared with respect to the chosen approach and methodology as well as results and clinical impact. Phenotypes analysed included tumours, diabetes mellitus, and cardiovascular diseases. All studies applied one or more statistical measures reporting on calibration, discrimination, or reclassification to quantify the benefit of including SNPs, but differed substantially regarding the methodological details that were reported. Several examples for improved risk assessments by considering disease-related SNPs were identified. Although the add-on benefit of including SNP genotyping data was mostly moderate, the strategy can be of clinical relevance and may, when being paralleled by an even deeper understanding of disease-related genetics, further explain the development of enhanced predictive and diagnostic strategies for complex diseases.

For the Legascreen consortium.

Members of the Legascreen consortium are listed in Acknowledgments.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-016-1636-z) contains supplementary material, which is available to authorized users.

✉ Bent Müller
bent.mueller@izi.fraunhofer.de

¹ Department of Cell Therapy, Fraunhofer Institute for Cell Therapy and Immunology, Perlickstraße 1, Leipzig, Germany

² Translational Center for Regenerative Medicine, Leipzig, Germany

³ Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Munich, Germany

⁴ Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

⁵ Institute for Human Genetics, University Hospital Leipzig, University of Leipzig, Leipzig, Germany

⁶ Institute for Human Genetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

⁷ Massachusetts General Hospital and Harvard Medical School, Stroke and Neurovascular Regulation Laboratory, Charlestown, MA, USA

⁸ Fraunhofer Research Institution of Marine Biotechnology and Institute for Medical and Marine Biotechnology, University of Lübeck, Lübeck, Germany

⁹ Institute for Medical Informatics, Statistics and Epidemiology and LIFE-Leipzig Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany

Introduction

Most human diseases and disorders result from a complex interplay between multiple genetic and environmental factors (Lander and Schork 1994). These conditions are commonly called complex diseases or disorders. Particularly when facing severe complex conditions, prevention medicine and the development of long-term curative strategies demand effective and reliable disease prediction. This, however, remains challenging.

This limitation may be at least partially due to the fact that the vast majority of standard disease prediction models omit genetic information. Instead, they solely rely on typical risk factors (hereinafter termed ‘traditional risk factors’) such as environmental exposures and intermediate phenotypes. The latter are defined as disease-related clinical or molecular measures that are related to the pathomechanism(s) underlying the disease of interest. Well-known examples for such traditional risk factors include a high body mass index (BMI) and high blood cholesterol in cardiovascular diseases (Yusuf et al. 2004).

On the contrary, the recent advancements in the field of complex disease genetics have paved the way for including genetic data in disease prediction models. Moreover, genotyping disease-specific genetic variants can be conducted independently of the tested individual’s age and is increasingly being considered an affordable routine diagnostic procedure. Although identified genetic risk variants explain only a minor proportion of heritability so far, this proportion is continually growing due to ongoing advances provided by genome-wide association studies (GWAS) and next generation sequencing analyses (Stranger et al. 2011). Particularly, GWAS have identified a growing number of common single nucleotide polymorphisms (SNPs) and several studies have started to consider such genetic information in the framework of common complex disease prediction with notable, but highly varying success. Here, we provide a systematic analysis of these studies by discussing the applied methodology, reliability of obtained results and their clinical relevance: based on this analysis, we further suggest potential directions for future research. We extend previous work (Cook and Paynter 2010; Thanassoulis and Vasan 2010; Wang 2011; Vassy and Meigs 2012) by including current original publications as well as by comparing results across prediction of different phenotypes. This allows us to analyse on a broader basis for key-drivers that may be related to improved prediction performance. Investigated parameters include the performance of the baseline model without genetics, the number of SNPs, the SNP validation level and that of the model, family history, and whether SNPs were chosen

that are associated with the predicted phenotype itself or associated with intermediate phenotypes. Based on this analysis, we further suggest potential directions for future research.

Search strategy and study identification

For direct comparison, we included studies that predicted susceptibility to frequent complex diseases and disorders by models incorporating (1) traditional (non-genetic) risk factors and (2) traditional risk factors and common genetic variants. Studies predicting the course of a disease were not considered. Moreover, studies selected had to test the benefit of genetic marker inclusion by comparing the combined prediction model against a model omitting genetic markers in a quantitative way. If any of these criteria was not met, the study was not enclosed.

The included studies are selected according to the following search strategy: initially, PubMed was filtered by the search term “Risk”[MeSH] AND “Genetic Predisposition to Disease”[MeSH] AND “Polymorphism, Single Nucleotide”[MeSH] AND improve*. Articles published between 2006/01/01 and 2015/10/01 were included, results were restricted to the species humans. Articles of the category ‘review’ and ‘clinical study’ were excluded. This resulted in 250 articles. When filtering these articles according to our inclusion criteria, 25 eligible studies remained. Subsequently, we iteratively analysed the co-citation network based on the reference lists of the included studies. This identified further 17 additional articles. Finally, another 13 articles were identified using the search engines <https://www.google.de/> and <http://scholar.google.de/> by combining the terms ‘SNPs’, ‘family information’ and ‘risk prediction’ (Fig. 1). With this strategy, we benefit on the one hand from the controlled MeSH vocabulary of MEDLINE which decreases loss of potentially relevant articles due to differences in vocabulary. On the other hand, analysis of the co-citation network extends this initial MEDLINE search in an expert-guided manner.

In total, we identified 55 studies meeting all pre-set inclusion criteria. Several studies analysed more than one cohort or used different analytic models. When multiple nested non-genetic models were reported for the same cohort, the best-performing model was preferred, which typically included the highest number of predictors. This resulted in 100 included distinct analyses (Table 1; Supplemental Table 1). Predicted phenotypes comprised cardiovascular pathologies ($n = 18$ studies) (Humphries et al. 2007; Morrison et al. 2007; Kathiresan et al. 2008; Paynter et al. 2009, 2010; Davies et al. 2010; Ripatti et al. 2010; Hughes et al. 2012; Lluís-Ganella et al. 2012; Hernesniemi et al. 2012;

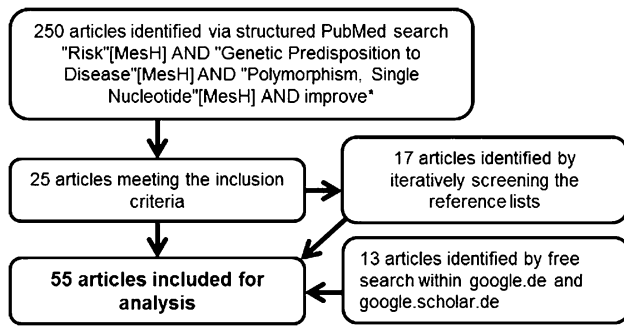


Fig. 1 Search strategy for the inclusion of studies in the analysis. This figure provides an overview of the search strategy and the numbers of eligible studies meeting the inclusion criteria

Brautbar et al. 2012; Isaacs et al. 2013; Bolton et al. 2013; Ganna et al. 2013; Tikkanen et al. 2013; Ibrahim-Verbaas et al. 2014; Beaney et al. 2015; de Vries et al. 2015), breast cancer ($n = 5$) (Wacholder et al. 2010; Mealiffe et al. 2010; Darabi et al. 2012; Dite et al. 2013; Vachon et al. 2015), prostate cancer ($n = 10$) (Zheng et al. 2008; Nam et al. 2009; Salinas et al. 2009; Aly et al. 2011; Johansson et al. 2012; Kader et al. 2012; Klein et al. 2012; Lindström et al. 2012; Helfand et al. 2013; Butoescu et al. 2014), type 2 diabetes ($n = 15$) (Balkau et al. 2008; van Hoek et al. 2008; Lysenko et al. 2008; Meigs et al. 2008; Lin et al. 2009; Schulze et al. 2009; Talmud et al. 2010; Wang et al. 2010; de Miguel-Yanes et al. 2011; Vassy et al. 2012a, b; Tam et al. 2013; Mühlenbruch et al. 2013; Vassy et al. 2014; Walford et al. 2014), atrial fibrillation ($n = 2$) (Everett et al. 2013; Tada et al. 2014), venous thrombosis ($n = 2$) (de Haan et al. 2012; Bruzelius et al. 2014), esophageal squamous cell carcinoma (ESCC) (Chang et al. 2013), melanoma (Fang et al. 2013) and Parkinson's disease (Hall et al. 2013) ($n = 1$ each). If not reported in the original publications, p values were calculated from reported confidence intervals.

Which genetic markers were selected to improve prediction?

We did not identify studies which conducted de novo SNP selections, but instead referred to previously published GWAS/candidate gene association studies which identified genetic variants associated with the disease of interest. Hence, SNPs used in the prediction studies could be considered pre-validated at least at a basic level. However, different SNP selection strategies were reported (Table 1). 48 studies (87 %) included SNPs resulting from previous GWAS while 7 studies (13 %) considered SNPs identified in previous candidate association studies. The method of choice strongly correlated with the year of publication, probably indicating an increased availability of GWAS data for the disease of

interest: only 2 out of 42 studies (5 %) published after 2009 used SNPs from candidate association studies. Contrarily, 5 out of 13 studies (39 %) published before 2010 solely relied on candidate association studies for SNP selection.

Notably, seven studies (Kathiresan et al. 2008; Paynter et al. 2010; Lluís-Ganella et al. 2012; Hernesniemi et al. 2012; Brautbar et al. 2012; Isaacs et al. 2013; Ibrahim-Verbaas et al. 2014) explicitly distinguished between SNPs that are directly associated with the predicted endpoint, and SNPs that are associated with intermediate phenotypes and therefore may contribute indirectly. A rationale behind this approach is that it may be of relevance for the predictive value of genetics when SNPs associated with intermediate phenotype are excluded as these intermediate phenotypes may be already included as predictors in the baseline model. A different rationale is that the model may benefit from shared genetics of the intermediate and predicted phenotype.

How were SNPs included into the prediction model?

In general, two main strategies for including genetic data into the predictive model were identified (see also Fig. 2). In the first, each genetic risk variant was considered as an individual covariate in addition to the traditional risk factors in the framework of a regression model. This strategy was adopted by 6/55 studies (11 %) (Paynter et al. 2009; Lindström et al. 2012; Chang et al. 2013; Bolton et al. 2013; Helfand et al. 2013; Bruzelius et al. 2014). A disadvantage of this approach is that the model size considerably increases with the number of independent genetic risk factors added. The second strategy aimed to solve this problem by proposing an additive genetic risk score (Horne et al. 2005), which was adopted by 49/55 studies (89 %). The simplest form to construct a genetic risk score is to count the number of risk alleles among all SNPs, which was performed in the majority of the studies (26/49 studies, 53 %). However, this assumes that each risk allele of each SNP has the same predictive value. However, this is likely to be divergent from reality in most cases. To account for potential differences between SNPs, SNPs can be alternatively weighted according to their effect size. This was done in 29/49 studies (53 %). Please note that studies utilising both strategies were assigned to both categories. Using an additive genetic risk score composed of weighted risk variants resulted in risk estimates with similar effect sizes as compared to risk estimates from a regression model including each genetic variant as an individual covariate. In the former approach, however, fewer degrees of freedom are utilised. The analysed studies adopted two methods for SNP weighting. In the first method, weights were based on effect sizes from an independent cohort, e.g. from the literature (25/29, 86 %). In the second, weights were calculated from the same population used

Table 1 Overview of the number and type of genetic data used for prediction

Predicted phenotype	Study	Type of genetic variants used
Breast cancer	Mealiffe et al. (2010)	7 GWAS-SNPs with independent replication
	Wacholder et al. (2010)	10 GWAS-SNPs
	Dite et al. (2013)	7 GWAS-SNPs with independent replication
	Darabi et al. (2012)	18 GWAS-SNPs
	Vachon et al. (2015)	76 GWAS-SNPs
Prostate cancer	Zheng et al. (2008)	5 SNPs from candidate association studies
	Nam et al. (2009)	19 SNPs from candidate association studies
	Salinas et al. (2009)	5 SNPs from GWAS/candidate studies showing independent cumulative association
	Aly et al. (2011)	35 GWAS-SNPs with independent replication
	Johansson et al. (2012)	33 GWAS-SNPs
	Kader et al. (2012)	33 GWAS-SNPs with independent replication
	Klein et al. (2012)	49 SNPs associated with prostate cancer or PS/1 SNP associated with breast cancer
	Lindström et al. (2012) (five age groups)	25 GWAS-SNPs
	Helfand et al. (2013)	4 SNPs from candidate association studies replicated in GWAS
	Butoescu et al. (2014)	9 GWAS-SNPs
Esophageal squamous cell carcinoma (ESCC)	Chang et al. (2013)	25 GWAS-SNPs
Melanoma	Fang et al. (2013)	11 GWAS-SNPs
Cardiovascular disease (CVD) related events	Humphries et al. (2007)	4 SNPs and 7 SNPs with interaction terms from candidate association studies
	Morrison et al. (2007)	10 SNPs for whites and 11 SNPs for blacks from candidate association studies
	Kathiresan et al. (2008)	9 GWAS-SNPs (associated with LDL and HDL)
	Paynter et al. (2009)	1 SNP from a candidate association studies with independent replication
	Davies et al. (2010)	13 GWAS-SNPs
	Paynter et al. (2010)	101 GWAS-SNPs associated with CVD or an intermediate phenotype and 12 GWAS-SNPs associated with CVD but no intermediate phenotype
	Ripatti et al. (2010)	13 GWAS-SNPs
	Brautbar et al. (2012)	13 SNPs associated with CVD but no intermediate phenotype from GWAS/candidate studies with independent replication
	Hernesniemi et al. (2012)	24 GWAS-SNPs (associated with CAD)
	Hughes et al. (2012)	11 SNPs and 2 haplotype from GWAS with independent replication and 15 SNPs from GWAS with independent replication
	Lluis-Ganella et al. (2012) (two cohorts)	8 GWAS-SNPs associated with CVD but not with intermediate phenotypes
	Bolton et al. (2013)	27 SNPs from meta-analysis of GWAS-SNPs
	Ganna et al. (2013)	395 SNPs (associated with CHD and intermediate phenotypes) from GWAS and 46 SNPs (directly associated with CHD) from GWAS
	Isaacs et al. (2013)	95 SNPs from meta-GWAS on TC, LDL-C, HDL-C, and TG
	Tikkanen et al. (2013)	28 SNPs from GWAS
	Ibrahim-Verbaas et al. (2014)	324 GWAS-SNPs (associated with all stroke risk domains)

Table 1 continued

Predicted phenotype	Study	Type of genetic variants used
Atrial fibrillation	Beaney et al. (2015)	13 GWAS-SNPs from CARDIoGRAMplusC4D and 19 SNPs from GWAS and candidate studies
	de Vries et al. (2015)	49 genome-wide significant SNPs and 103 SNPs at FDR <10 %
	Everett et al. (2013)	12 GWAS-SNPs
Venous thrombosis	Tada et al. (2014)	12 GWAS-SNPs
	de Haan et al. (2012)	5 SNPs from GWAS/candidate studies with independent replication
Type 2 diabetes	Bruzelius et al. (2014)	7 SNPs from GWAS and candidate association studies
	Balkau et al. (2008)	2 SNPs from candidate association studies
	Lyssenko et al. (2008)	11 GWAS-SNPs
	Meigs et al. (2008)	18 GWAS-SNPs
	van Hoek et al. (2008)	18 GWAS-SNPs with independent replication
	Lin et al. (2009)	15 GWAS-SNPs
	Schulze et al. (2009)	20 GWAS-SNPs
	Talmud et al. (2010)	20 GWAS-SNPs with independent replication
	Wang et al. (2010)	19 SNPs from candidate studies with independent replication
	de Miguel-Yanes et al. (2011)	40 GWAS-SNPs
	Vassy et al. (2012a)	38 GWAS-SNPs
	Vassy et al. (2012b)	38 GWAS-SNPs
	Mühlenbruch et al. (2013)	42 GWAS-SNPs
	Tam et al. (2013)	14 GWAS-SNPs
Vassy et al. (2014)	62 GWAS-SNPs	
Walford et al. (2014)	62 GWAS-SNPs	
Parkinson disease	Hall et al. (2013)	4 GWAS-SNPs

“GWAS-SNPs” were identified in previous genome-wide association studies. In contrast, “SNPs” were derived from previous locus-wide association studies

“SNPs with independent replication” are defined as SNPs identified in a previous study, replicated in another previous association study, and finally used in the prediction study. Otherwise, SNPs resulted from at least one independent discovery study

for prediction analysis (6/29, 14 %), which potentially might result in biased estimates of the model’s prediction accuracy.

Next to these two commonly applied strategies for SNP inclusion into the prediction model, two additional approaches were identified. Humphries et al. (2007) focused on the independent component of each SNP and, given all traditional risk factors, on the effect of all other SNPs. By doing so, the authors adjusted the effect of each SNP on all other SNPs and known risk factors, and extended the traditional risk score using these adjusted values. Helfand et al. (2013) did not include genetic variants in their model, but instead aimed to improve the value of prostate specific antigen (PSA) levels as an indicator for biopsy-based screening of prostate cancer with genetic data. The authors divided the PSA level by a genetic risk score and used resulting, modified PSA levels to decide on recommendation for biopsy. This represents a rather unconventional approach which has only rarely been reported so far. Hence, additional in-depth assessment of its impact and predictive value is recommendable.

Regardless of the outlined differences, many studies accounted for the correlation of individual SNPs, which is a result from linkage disequilibrium. The common strategy was pruning SNPs, i.e. removing one SNP of each correlating pair above a certain cut-off. Another approach to account for the correlation between SNPs would be to explicitly model their correlation structure. However, this was not done in any of the investigated studies. We note that methodology for feature selection and predictor weighting is continually improving (Kooperberg et al. 2010; Kruppa et al. 2012) and future work will show whether this may lead to further improved prediction.

How were traditional risk factors included into the prediction model?

Traditional risk factors were incorporated in the prediction models in different ways. On the one hand, studies

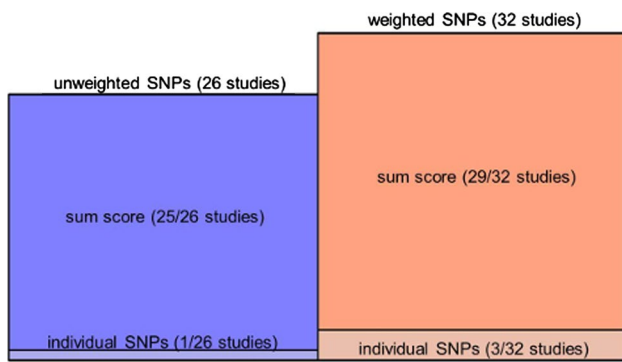


Fig. 2 Overview of methods as to how genetic data were included in the prediction model. “Sum score”: from all SNPs a single predictor reflecting the genetic burden was created and used as single parameter in the prediction model, “individual SNPs”: SNPs were included as individual covariates in the model used for prediction, “weighted”: risk alleles of SNPs were weighted according to the respective odds ratio, and “unweighted”: risk alleles of SNPs were counted without weighting. Note that Brautbar et al. (2012), Everett et al. (2013) and Talmud et al. (2010) used weighted as well as unweighted sum scores in their analyses and thus appear in both categories in the figure

used well-established risk models like the Framingham risk score for cardiovascular disease (e.g. Isaacs et al. 2013; Bolton et al. 2013) or the Gail model for breast cancer (e.g. Wacholder et al. 2010; Darabi et al. 2012). On the other hand, some studies selected non-genetic risk factors themselves for prediction. Ten of eighteen (56 %) studies predicting cardiovascular disease used established risk models as well as all (5/5) studies predicting breast cancer. When selecting non-genetic risk factors themselves, overfitting might occur if no appropriate validation strategy is applied.

Note that heterogeneity among prediction performance of the baseline model can be high between different cohorts and studies, even when the same model of traditional risk factors is applied. Exemplarily, the AUC for predicting prostate cancer with the non-genetic PCPTRC risk model ranged 0.56–0.72 when it was applied to different cohorts in the same study (Ankerst et al. 2012). Therefore, improvements of prediction due to genetics should always be interpreted under consideration of this heterogeneity (Ankerst and Thompson 2012).

How can the benefit of genetic data inclusion for disease prediction be measured?

Several methods have been reported to compare models including genetic data with those omitting such information. Most of these methods can be categorised in discrimination ability, reclassification ability, and model calibration (Steyerberg et al. 2010; Pencina et al. 2010; Wang 2011; Siontis et al. 2012).

Discrimination describes the ability of a model to distinguish individuals from risk and non-risk groups. The most prominent measure is the area under the receiver operating characteristic (ROC) curve (AUC). Within a classical ROC, the sensitivity of the model is plotted against 1-specificity for various thresholds of the predictive score (Bamber 1975). The AUCs can range from 0.5 (the model is completely uninformative) to 1 (perfect discrimination between affected and unaffected individuals). Technically speaking, the AUC can be interpreted as the probability that an affected individual has a higher predicted risk score than an individual from the control group (Hanley and McNeil 1982). In order to analyse the benefit of including genetic data in a disease prediction model, the AUC of the model additionally including genetic data is compared with the AUC of the baseline model. If the former is significantly larger, a benefit can be claimed. However, the AUC under the ROC method has certain limitations. First, it represents a measure of multiple realisations of a predictive model as it evaluates the performance of the model for all possible thresholds of the predictive score. This is done regardless of whether or not these thresholds are clinically meaningful. Therefore, an improved AUC does not necessarily reflect an improved performance with respect to clinical relevance. Second, the AUC has been criticised for being relatively insensitive (Cook 2007). This is of particular relevance when the AUC of the baseline model is already good. Here, the power to detect a statistically significant improvement of the AUC by including a certain genetic marker is much lower than the power to improve the AUC of a model with lower initial AUC values (Tzoulaki et al. 2009; Pencina et al. 2010). A prominent example is the study of Pencina et al. (2010), who reported that an additional genetic marker with an effect size of 0.41 (corresponding to an SNP with an odds ratio of 1.5) can improve the AUC of baseline models with values 0.55, 0.6, and 0.75 to an AUC of 0.63, 0.65, and 0.77, respectively. Such calculations typically do not assume a correlation between the predictors of the baseline model and the predictors of the model including genetic data. As one may expect, adding a genetic component to a baseline model will result in a relevant correlation between predictors of both models. Hence, the power of the AUC method may be even lower.

Reclassification improvement describes improvement in the classification of cases and controls when comparing an updated model against a baseline model. For this purpose, Pencina et al. (2010) proposed the net reclassification improvement (NRI). The NRI examines whether the model including genetic data shifts cases to higher risk categories more often than to lower risk and, vice versa, controls to lower risk categories more often than to higher ones. Improvement can be claimed if the sum of these

movements is better than 0 %, which is found if there is equal movement in the correct and incorrect direction (Paynter et al. 2010). However, changes of risk categories do not necessarily result from clinically important risk categories. Therefore, some authors report the ‘clinical NRI’ related to changes within the most clinically relevant risk categories (Cook and Paynter 2011). Although the NRI is a well-suited measure for reclassification analysis, a number of limitations need to be considered (see Cook and Paynter 2011 for details). Therefore, the results obtained should be interpreted carefully and reported in sufficient detail (Pepe 2011).

The integrated discrimination improvement (IDI) represents an alternative and relevant reclassification measure. The IDI is the difference between the discrimination slope of the baseline and the updated model (Pencina et al. 2008). As clinical risk categories are not required for IDI calculation, it is of particular value when such categories do not (yet) exist.

Calibration assesses the agreement of predicted and observed risks across subgroups with varying baseline risk. In general, only predicted risks that are well-calibrated are useful for clinical management, because treatment decisions often depend on estimates of the predicted risk. The most common measure for calibration is the Hosmer–Lemeshow test, which compares predicted and observed outcomes over percentiles of risk (Lemeshow and Hosmer 1982). Superiority of one model to another is typically demonstrated by increased p values resulting from this test.

All measures described above reflect different aspects of model quality and should be considered in close relation to each other whenever possible. For example, a study in which the AUC increase is considered small may still provide substantial improvement of the reclassification measure NRI and/or increase in the IDI (Pencina et al. 2008). Therefore, an AUC increase of even 0.01 might still be suggestive of a meaningful improvement in some cases (Pencina et al. 2008), as reclassification of clinically important patient subgroups might have been improved. However, reclassification and discrimination are only of clinical value when the predicted risks are in strong correlation with the actual risk. The estimation of calibration is therefore necessary, and, in the case of a poorly calibrated model, a recalibration to the population of interest is strongly recommended (Pepe and Janes 2013).

In order to avoid biased estimates, the quality measures discussed should be computed on test sets independent from the initial (‘training’) set used for fitting the model. All prediction studies considered in this review only included SNPs that were pre-validated in previous, independent association studies.

Did the inclusion of genetic information improve disease prediction?

We identified both, studies reporting and not reporting improved prediction when including genetic data (Supplemental Table 1; Fig. 3). Thirty studies (55 %) saw significant improvements in AUC when including genetic data. Fourteen further studies (26 %) did not identify a significant AUC improvement, but a significant improvement in reclassification. Effect sizes of traditional risk factors were generally larger than those of genetic risk factors, regardless of whether or not they were determined in an independent data set. Nevertheless, a considerable variation in the effect sizes of traditional risk factors was observed. For example, a strong risk factor for venous thrombosis is the presence of minor leg injuries. This risk factor has an odds ratio (OR) >5 (Previtali et al. 2011), whereas obesity is a moderate risk factor reported to confer risk for cardiovascular diseases with an OR of 1.62 (Yusuf et al. 2004). As the power to improve the AUC in prediction depends on the predictive strength of the baseline model, poorly performing baseline models showed best improvement when adding genetic predictors. In consequence, we found a clear relationship between the improvement of the AUC due to including genetic data and the phenotypes predicted (Fig. 3). For example, Zheng et al. (2008) (predicting prostate cancer) reported an AUC of 0.608 for a model accounting for age, geographic region and family history. After the addition of the genetic risk score, the AUC of the model increased with statistical significance to 0.633 ($p = 6.1 \times 10^{-6}$). In well-performing baseline models, a significant increase of the AUC is less frequently reported, still, prediction improvement by considering genetic data is often shown by applying measures of reclassification. As an example, Walford et al. (2014) reported no significant improvement of the baseline model AUC (AUC = 0.861), but significant improvement in the reclassification measure (NRI = 0.247, $p = 0.0009$).

Of note, study comparability was limited by divergent study aims and designs as well as by the different analytic strategies applied. For example, four studies did not analyse whether the observed changes in discrimination were statistically significant. 15/55 studies (27 %) reported results of discrimination, reclassification, and calibration of the model with and without genetic data conjointly (Supplemental Table 1). Furthermore, relevant aspects of statistical procedures were frequently not reported in detail. For example, studies applying bootstrap-based procedures rarely provided details whether weighting of SNPs or assessment of predictive accuracy was done using the out-of-bag or the in-bag data. As another example, studies applying cross-validation rarely provided details whether weighting of SNPs or traditional risk factors was done

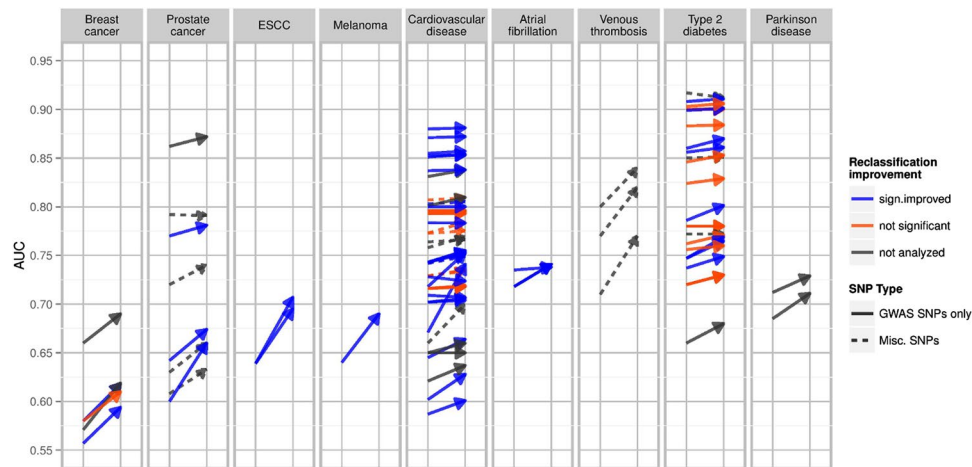


Fig. 3 Overview of the discrimination improvement due to inclusion of genetic data across all included 100 analyses. An AUC of 1.0 indicates perfect discrimination between cases and controls, 0.5 is equivalent to random guessing. Studies are stratified according to their predicted phenotype. Each reported analysis is depicted in form of an *arrow* with the *arrow* start indicating the AUC when using traditional risk factors only and the *arrowhead* indicating the AUC of the model including genetic data. The *colour* of the *arrow* illustrates significance of reclassification measures with *blue* statistically sig-

nificant, *orange* not statistically significant, and *grey* not tested. *Solid lines* indicate GWAS-derived SNPs and *dashed lines* all other SNPs. The figure clearly illustrates it is generally harder to improve discrimination of a prediction model by including the genetic data in cases where the baseline model already performs well. Nevertheless, in some cases significant reclassification can be observed even for high baseline AUC values. For numbers and additional details on studies, please also refer to Supplemental Table 1

only once before the cross-validation procedure started or repeatedly within each cross-validation iteration. Such details are very helpful when evaluating reported classification performance and are important for a valid comparison of results from different studies. Guidelines already exists, but are only infrequently accounted for in current publications (Janssens et al. 2011).

What is characteristic for studies showing the strongest improvement of prediction by the inclusion of genetic data?

Several studies report enhanced prediction although performance of the baseline model was similar to other studies not reporting such improvement (Fig. 3). The first outstanding example is the study of Bolton et al. (2013) predicting coronary heart disease. Here, AUC increased in two analyses from 0.671 to 0.741 and from 0.717 to 0.753 when including genetic data. Although this study has several strengths (large number of recently reported SNPs, a well-defined phenotype definition, and a prospective, population-based design) it also comes with some limitations. First, the sample size is rather moderate compared to other studies on the same phenotype. Second, each SNP was included in the model without applying weights from the literature, but estimating them from the cohort also used for prediction analysis. This may result in biased estimates. Third, the baseline model was one of the weakest

that existed for this particular phenotype, which clearly favoured significant prediction improvement by adding a genetic component to the model. Other interesting examples are two studies predicting the occurrence of venous thrombosis (de Haan et al. 2012; Bruzelius et al. 2014). Here, AUC increased from 0.77 to 0.82 and from 0.71 to 0.77 in the discovery and validation cohort, respectively (de Haan et al. 2012) and Bruzelius et al. (2014) reported an increased AUC from 0.80 to 0.84. These studies, in difference to all others investigated, included common SNPs with very strong effect sizes: variants rs6025 and rs8176719 have literature-reported ORs of 3.8 and 1.85 with a frequency in cases of 10 and 47 %, respectively. Such impressive effect sizes together with high minor allele frequencies are of course the exception for genetic factors of common diseases. If present, however, they allow for a tremendous improvement of prediction when added.

Are there SNP-specific differences?

In addition to the effect sizes, we explored whether additional SNP characteristics may lead to better disease prediction. First, we compared performance of GWAS SNPs versus SNPs from candidate studies. 32/48 (67 %) studies that included SNPs from GWAS reported significant improvement in classification, 18/32 (56 %) also included some strategy of validation. 4/7 (57 %) studies that

included SNPs from candidate association studies reported significant improvement in classification, with two studies including a validation strategy. A reason for non-superior performance of GWAS SNPs might be that candidate SNPs were well chosen focussing on well-validated SNPs with at least moderate effect sizes.

Analyses applying weighted genetic risk scores more frequently reported a significant improvement in prediction due to genetic data (40/58, 69 %) as compared to analyses applying non-weighted genetic risk scores (17/37, 46 %). An improved performance compared with studies applying non-weighted genetic risk scores was still observed when we filtered for analyses applying SNP weights determined in independent cohorts (29/43, 67 %). This was also true when further filtering for studies completely evaluating in a second cohort (4/4, 100 %). Therefore, it can be assumed that better performance of weighted genetic risk scores is unlikely to solely result from model overfitting. It also underpins the fact that weights from independent cohorts should be used to maximise the reproducibility of results whenever possible.

Inclusion of familial risk did not have a major effect on the predictive power of candidate SNPs. 30/50 (60 %) analyses with familial risk had improvement of prediction when including genetic data versus 32/50 (64 %) of analyses that did not include familial risk. This finding is in accordance with previous reports (Ripatti et al. 2010; Vassy and Meigs 2012). As noted by Ripatti et al. (2010), reasons include measurement error for family history and that currently known genetic variants only account for a small proportion of familial risk.

Finally, we did not observe a major effect of number of SNPs on prediction improvement (Supplemental Fig. 1). This may reflect the limited knowledge regarding the genetic component of complex diseases. Simulations demonstrate that the prediction improvement might be considerably higher when all relevant SNPs were included (Aly et al. 2011; Dudbridge 2013), but considerably larger studies are needed to identify more relevant SNPs and to uncover a larger fraction of the heritability.

Is there a benefit of including or excluding SNPs associated with intermediate phenotypes?

Brautbar et al. (2012), Paynter et al. (2010) and Lluís-Ganella et al. (2012) investigated the strategy of including SNPs significantly associated with coronary heart disease, but excluding SNPs associated with an intermediate phenotype related to this disease. In contrast, Kathiresan et al. (2008) and Isaacs et al. (2013) restricted selected SNPs to those that were associated with an intermediate phenotype. Ibrahim-Verbaas et al. (2014) included 322 SNPs associated

with nine intermediate phenotypes of stroke and only two SNPs directly associating with stroke. Only Paynter et al. (2010) directly compared both strategies in the same dataset and found a similar (weak) performance of SNPs from both SNP selection strategies. Brautbar et al. (2012), Kathiresan et al. (2008), Lluís-Ganella et al. (2012) and Ibrahim-Verbaas et al. (2014) observed an improved predictive value when including genetic information. Given those examples for a successful inclusion as well as exclusion of SNPs associated with intermediate phenotypes, no final conclusion can be drawn at this point due to the limited number of studies directly comparing different strategies at the same cohort. However, investigations of endophenotypes appear to increase in relevance while applied methodology continually improves. Significant benefit is expected especially for complex diseases in which endophenotypes have the potential to bridge complex genetic backgrounds and known disease heterogeneity (Insel and Cuthbert 2009). Interestingly, a complementary strategy applied by Hernesniemi et al. (2012) proved to be ineffective. Selecting GWAS-derived SNPs associated with cardiovascular diseases (CVD) to predict an intermediate phenotype of CVD, i.e. intima-media thickness and artery elasticity, did not result in enhanced prediction while only a limited association with the intermediate phenotype was found. A plausible explanation is that selected SNPs may act via different intermediate phenotypes.

What are potential implications for clinical research and practice?

The inclusion of genetic markers such as SNPs into diagnostic procedures was originally thought to rapidly increase diagnostic accuracy and to significantly reduce the number of patients being diagnosed false-negatively (Ginsburg and McCarthy 2001; Diamandis et al. 2010). In theory, this would ultimately translate into both faster and improved therapeutic intervention and preventive treatment. After analysing results from first studies, much of the initial enthusiasm cooled off. Ripatti et al. (2010) and de Haan et al. (2012) were sceptical about the potential clinical use of common SNPs for disease prediction in cardiovascular diseases. Wacholder et al. (2010) found a certain benefit in prediction when including genetic markers for high- and low-risk groups, but not in groups at intermediate risk for breast cancer (2010). Still, the benefit was not sufficiently large to meaningfully improve the identification of patients who might profit from prophylactic treatment. Nevertheless, a clear diagnostic advantage was repeatedly seen for a number of diseases when genetic markers for diagnosis were included (Fig. 3), suggesting the definition of particular scenarios under which this strategy can provide a measurable benefit.

- Power is a key-driver whether or not the genetic test component improved accuracy of prediction. The better the baseline model performs, the harder it is to improve. Therefore, the careful assessment of the baseline model performance is highly relevant to estimate how much improvement can be realistically expected by adding genetic information.
- Risk prediction studies should seek to avoid bias, preferably by performing model selection and evaluation in independent training and testing cohorts, or at least by applying resampling methods (Steyerberg et al. 2001).
- Genetic variants included in the model should be comprehensive and well validated, i.e. confirmed by additional studies investigating different cohorts. Moreover, they should be regularly updated if new genetic associations are reported.
- More work is required to unravel how the inclusion of SNPs relevant for intermediate phenotypes might improve prediction of the disease phenotype.
- The majority of scenarios might benefit from the use of weighted genetic risk variants in the prediction model instead of using non-weighted counts of risk alleles.
- Several complementary quality measures that describe model performance should be reported for each approach. As a recommendation, this might include AUC, reclassification measures, recalibration, and measures of clinical relevance such as the NNS. This will facilitate the interpretation of results as well as the comparison of different studies.

Fig. 4 Take-home messages for predicting complex diseases with common genetic markers

From a socio-economic perspective, cost-effectiveness is relevant when deciding which patient to test. For example, Mealiffe et al. (2010) proposed to improve the cost-effectiveness by only testing patients whose risk status is likely to change. Only testing individuals close to classification cut-offs would lower the effort because the majority of patients do not require screening. However, with the steady progress in the field, the costs for thorough genetic profiling are continuously declining (Gershon et al. 2011). This in turn may increase the value of genetic diagnostics even for conditions where current approaches only provide a small improvement.

More recent studies including those by Ganna et al. (2013) and Tikkanen et al. (2013) are more optimistic about the benefit of adding genetic markers to disease prediction models. Ganna et al. (2013) estimated that one additional event resulting from coronary heart disease could be prevented in 318 patients by including genetic markers to their model. This measure is the gain in ‘number needed to screen’ (NNS) when comparing the baseline model with the model including genetic data. NNS is defined as the number of patients needed to be screened to detect one affected individual (Rembold 1998). Tikkanen et al. (2013) reported a gain in NNS of 15.9 (prevention of 135 events out of 2144 cases) by additional genetic testing in a subgroup of patients. This subgroup was previously categorised at

intermediate risk for coronary heart disease by using non-genetic information. While these findings are generally encouraging, the reported gains in NNS are still relatively high. Several strategies may be suggested to compensate for this problem and to ensure a measurable benefit.

First, reporting the NNS is generally recommended when investigating diagnostic procedures amended by genetic test components. This is important, since the NNS is rarely reported in the literature, although this measure precisely indicates clinical relevance and assesses the value of intervention. Consequently, the NNS can be used for identifying the best-performing model when comparing different approaches.

Second, future research may also focus on assessments in which the investigation of traditional risk factors is more costly than the investigation of genetic risk factors. By applying a second confirmative diagnosis, e.g. by conventional diagnostics using non-genetic components, the number of patients treated with optimal benefit may be maximised (Peterson et al. 2013; Hagemann et al. 2013).

Third, future work may focus on conditions for which a highly effective or even causative treatment is available and/or early therapeutic intervention provides a clear benefit with respect to prognosis. This may further include cases in which a failure to treat causes relevant additive burden to the patient. In these scenarios, prediction is of high

clinical relevance and even a small benefit as the inclusion of genetic data may be of value. However, such approaches always require careful balancing of socio-economic costs caused by additional treatments versus gained benefits, as well as a thorough ethical consideration.

Such improvement of diagnostic genetic markers might be also relevant for other fields of clinical research, e.g. adaptive clinical trials or improved handling of patient heterogeneity by molecular reclassification. Up to now, genetic testing for drug efficiency is rarely used in clinical practice albeit its underlying potential (Antman et al. 2012). This is also true for classifying subtyping of diseases using common SNPs. Here, promising candidates for certain diseases like breast cancer exist, but these are not routinely available yet (Garcia-Closas et al. 2013).

Conclusion

A considerable number of reports indicating that genetic data could contribute to the improvement of prediction models have been identified. The additive value of considering genetic information was statistically significant in many analyses, though limited with respect to the absolute effect in most cases. Hence, considerable progress is still required before routine clinical practice will benefit from including genetic data in the prediction of risk to certain diseases. Although the heterogeneity of the included phenotypes by the reviewed studies requires careful and case-specific interpretation, we derived general conclusions for future work which we summarised in Fig. 4.

Given encouraging examples of improved prediction with noticeable clinical relevance and in light of the ongoing progress in the field of genetics, we feel optimistic that including the genetic component in prediction models of complex diseases and disorders will continually increase and provide a measurable add-on benefit in the future.

Acknowledgments We would like to thank all members of the Legascreen consortium for their support during this study. Legascreen is funded by the Fraunhofer Society and the MaxPlanckSociety as a project within the framework of the “Pakt für Forschung und Innovation”. Dr. Holger Kirsten was also funded by the Leipzig Interdisciplinary Research Cluster of Genetic Factors, Clinical Phenotypes and Environment (LIFE Center, Universität Leipzig). The Legascreen consortium consists of: Prof. Dr. Dr. h.c. Angela D. Friederici, Dr. Jens Brauer, Dr. Michael Skeide, Dr. Nicole Neef, Gesa Schaadt, Indra Kraft, and Nadin Bobovnikov (all Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany) as well as Prof. Dr. Frank Emmrich, Dr. Arndt Wilcke, Dr. Holger Kirsten, Prof. Dr. Dr. Johannes Boltze, Bent Müller and Ivonne Czepezauer (Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany).

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Aly M, Wiklund F, Xu J et al (2011) Polygenic risk score improves prostate cancer risk prediction: Results from the Stockholm-1 cohort study. *Eur Urol* 60:21–28. doi:10.1016/j.eururo.2011.01.017
- Ankerst DP, Thompson IM (2012) Words of wisdom. Re: Combining 33 genetic variants with prostate-specific antigen for the prediction of prostate cancer: longitudinal study. *Eur Urol* 62:180. doi:10.1016/j.eururo.2012.04.010
- Ankerst DP, Boeck A, Freedland SJ et al (2012) Evaluating the PCPT risk calculator in ten international biopsy cohorts: results from the Prostate Biopsy Collaborative Group. *World J Urol* 30:181–187. doi:10.1007/s00345-011-0818-5
- Antman E, Weiss S, Loscalzo J (2012) Systems pharmacology, pharmacogenetics, and clinical trial design in network medicine. *Wiley Interdiscip Rev Syst Biol Med* 4:367–383. doi:10.1002/wsbm.1173
- Balkau B, Lange C, Fezeu L et al (2008) Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). *Diabetes Care* 31:2056–2061. doi:10.2337/dc08-0368
- Bamber D (1975) The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol* 12:387–415. doi:10.1016/0022-2496(75)90001-2
- Beaney KE, Cooper JA, Ullah Shahid S et al (2015) Clinical Utility of a Coronary Heart Disease Risk Prediction Gene Score in UK Healthy Middle Aged Men and in the Pakistani Population. *PLoS One* 10:e0130754. doi:10.1371/journal.pone.0130754
- Bolton JL, Stewart MCW, Wilson JF et al (2013) Improvement in prediction of coronary heart disease risk over conventional risk factors using SNPs identified in genome-wide association studies. *PLoS One* 8:e57310. doi:10.1371/journal.pone.0057310
- Brautbar A, Pompeii LA, Dehghan A et al (2012) A genetic risk score based on direct associations with coronary heart disease improves coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC), but not in the Rotterdam and Framingham Offspring, Studies. *Atherosclerosis* 223:421–426. doi:10.1016/j.atherosclerosis.2012.05.035
- Bruzelius M, Bottai M, Sabater-Lleal M et al (2014) Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost* 53:219–227. doi:10.1111/jth.12808
- Butoescu V, Ambroise J, Stainier A et al (2014) Does genotyping of risk-associated single nucleotide polymorphisms improve patient selection for prostate biopsy when combined with a prostate cancer risk calculator? *Prostate* 74:365–371. doi:10.1002/pros.22757
- Chang J, Huang Y, Wei L et al (2013) Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. *Carcinogenesis* 34:1782–1786. doi:10.1093/carcin/bgt106
- Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935. doi:10.1161/CIRCULATIONAHA.106.672402
- Cook NR, Paynter NP (2010) Genetics and breast cancer risk prediction—are we there yet? *J Natl Cancer Inst* 102:1605–1606. doi:10.1093/jnci/djq413
- Cook NR, Paynter NP (2011) Performance of reclassification statistics in comparing risk prediction models. *Biom J* 53:237–258. doi:10.1002/bimj.201000078
- Darabi H, Czene K, Zhao W et al (2012) Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Res* 14:R25. doi:10.1186/bcr3110

- Davies RW, Dandona S, Stewart AFR et al (2010) Improved prediction of cardiovascular disease based on a panel of single nucleotide polymorphisms identified through genome-wide association studies. *Circ Cardiovasc Genet* 3:468–474. doi:[10.1161/CIRCGENETICS.110.946269](https://doi.org/10.1161/CIRCGENETICS.110.946269)
- De Haan HG, Bezemer ID, Doggen CJM et al (2012) Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood* 120:656–663. doi:[10.1182/blood-2011-12-397752](https://doi.org/10.1182/blood-2011-12-397752)
- De Miguel-Yanes JM, Shrader P, Pencina MJ et al (2011) Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes Care* 34:121–125. doi:[10.2337/dc10-1265](https://doi.org/10.2337/dc10-1265)
- De Vries PS, Kavousi M, Ligthart S et al (2015) Incremental predictive value of 152 single nucleotide polymorphisms in the 10-year risk prediction of incident coronary heart disease: the Rotterdam Study. *Int J Epidemiol* 44:682–688. doi:[10.1093/ije/dyv070](https://doi.org/10.1093/ije/dyv070)
- Diamandis M, White NMA, Yousef GM (2010) Personalized medicine: marking a new epoch in cancer patient management. *Mol Cancer Res* 8:1175–1187. doi:[10.1158/1541-7786.MCR-10-0264](https://doi.org/10.1158/1541-7786.MCR-10-0264)
- Dite GS, Mahmoodi M, Bickerstaffe A et al (2013) Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model. *Breast Cancer Res Treat* 139:887–896. doi:[10.1007/s10549-013-2610-2](https://doi.org/10.1007/s10549-013-2610-2)
- Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 9:e1003348. doi:[10.1371/journal.pgen.1003348](https://doi.org/10.1371/journal.pgen.1003348)
- Everett BM, Cook NR, Conen D et al (2013) Novel genetic markers improve measures of atrial fibrillation risk prediction. *Eur Heart J* 34:2243–2251. doi:[10.1093/eurheartj/ehd033](https://doi.org/10.1093/eurheartj/ehd033)
- Fang S, Han J, Zhang M et al (2013) Joint effect of multiple common SNPs predicts melanoma susceptibility. *PLoS One* 8:1–9. doi:[10.1371/journal.pone.0085642](https://doi.org/10.1371/journal.pone.0085642)
- Ganna A, Magnusson PKE, Pedersen NL et al (2013) Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler Thromb Vasc Biol* 33:2267–2272. doi:[10.1161/ATVBAHA.113.301218](https://doi.org/10.1161/ATVBAHA.113.301218)
- Garcia-Closas M, Couch FJ, Lindstrom S, et al (2013) Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* 45:392–8, 398e1–2. doi: [10.1038/ng.2561](https://doi.org/10.1038/ng.2561)
- Gershon ES, Alliey-Rodríguez N, Liu C (2011) After GWAS: searching for genetic risk for schizophrenia and bipolar disorder. *Am J Psychiatry* 168:253–256. doi:[10.1176/appi.ajp.2010.10091340](https://doi.org/10.1176/appi.ajp.2010.10091340)
- Ginsburg GS, McCarthy JJ (2001) Personalized medicine: Revolutionizing drug discovery and patient care. *Trends Biotechnol* 19:491–496
- Hagemann IS, Cottrell CE, Lockwood CM (2013) Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Cancer Genet* 206:420–431. doi:[10.1016/j.cancergen.2013.11.003](https://doi.org/10.1016/j.cancergen.2013.11.003)
- Hall TO, Wan JY, Mata IF et al (2013) Risk prediction for complex diseases: application to Parkinson disease. *Genet Med* 15:361–367. doi:[10.1038/gim.2012.109](https://doi.org/10.1038/gim.2012.109)
- Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36
- Helfand BT, Loeb S, Hu Q et al (2013) Personalized prostate specific antigen testing using genetic variants may reduce unnecessary prostate biopsies. *J Urol* 189:1697–1701. doi:[10.1016/j.juro.2012.12.023](https://doi.org/10.1016/j.juro.2012.12.023)
- Hernesniemi JA, Seppälä I, Lyytikäinen L-P et al (2012) Genetic profiling using genome-wide significant coronary artery disease risk variants does not improve the prediction of subclinical atherosclerosis: the Cardiovascular Risk in Young Finns Study, the Bogalusa Heart Study and the Health 2000 Survey—a meta-analysis of three independent studies. *PLoS One* 7:e28931. doi:[10.1371/journal.pone.0028931](https://doi.org/10.1371/journal.pone.0028931)
- Horne BD, Anderson JL, Carlquist JF et al (2005) Generating genetic risk scores from intermediate phenotypes for use in association studies of clinically significant endpoints. *Ann Hum Genet* 69:176–186. doi:[10.1046/j.1529-8817.2005.00155.x](https://doi.org/10.1046/j.1529-8817.2005.00155.x)
- Hughes MF, Saarela O, Stritzke J et al (2012) Genetic markers enhance coronary risk prediction in men: the MORGAM prospective cohorts. *PLoS One* 7:e40922. doi:[10.1371/journal.pone.0040922](https://doi.org/10.1371/journal.pone.0040922)
- Humphries SE, Cooper JA, Talmud PJ, Miller GJ (2007) Candidate gene genotypes, along with conventional risk factor assessment, improve estimation of coronary heart disease risk in healthy UK men. *Clin Chem* 53:8–16. doi:[10.1373/clinchem.2006.074591](https://doi.org/10.1373/clinchem.2006.074591)
- Ibrahim-Verbaas CA, Fornage M, Bis JC et al (2014) Predicting stroke through genetic risk functions: the CHARGE Risk Score Project. *Stroke* 45:403–412. doi:[10.1161/STROKEAHA.113.003044](https://doi.org/10.1161/STROKEAHA.113.003044)
- Insel TR, Cuthbert BN (2009) Endophenotypes: bridging genomic complexity and disorder heterogeneity. *Biol Psychiatry* 66:988–989. doi:[10.1016/j.biopsych.2009.10.008](https://doi.org/10.1016/j.biopsych.2009.10.008)
- Isaacs A, Willems SM, Bos D et al (2013) Risk scores of common genetic variants for lipid levels influence atherosclerosis and incident coronary heart disease. *Arterioscler Thromb Vasc Biol* 33:2233–2239. doi:[10.1161/ATVBAHA.113.301236](https://doi.org/10.1161/ATVBAHA.113.301236)
- Janssens ACJW, Ioannidis JPA, Van Duijn CM et al (2011) Strengthening the reporting of genetic risk prediction studies: The GRIPS statement. *Eur J Clin Invest* 41:1004–1009. doi:[10.1111/j.1365-2362.2011.02494.x](https://doi.org/10.1111/j.1365-2362.2011.02494.x)
- Johansson M, Holmström B, Hinchliffe SR et al (2012) Combining 33 genetic variants with prostate-specific antigen for prediction of prostate cancer: longitudinal study. *Int J Cancer* 130:129–137. doi:[10.1002/ijc.25986](https://doi.org/10.1002/ijc.25986)
- Kader AK, Sun J, Reck BH et al (2012) Potential impact of adding genetic markers to clinical parameters in predicting prostate biopsy outcomes in men following an initial negative biopsy: Findings from the REDUCE trial. *Eur Urol* 62:953–961. doi:[10.1016/j.eururo.2012.05.006](https://doi.org/10.1016/j.eururo.2012.05.006)
- Kathiresan S, Melander O, Anevski D et al (2008) Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 358:1240–1249
- Klein RJ, Hallden C, Gupta A et al (2012) Evaluation of multiple risk-associated single nucleotide polymorphisms versus prostate-specific antigen at baseline to predict prostate cancer in unscreened men. *Eur Urol* 61:471–477. doi:[10.1016/j.eururo.2011.10.047](https://doi.org/10.1016/j.eururo.2011.10.047)
- Kooperberg C, LeBlanc M, Obenchain V (2010) Risk prediction using genome-wide association studies. *Genet Epidemiol* 34:643–652. doi:[10.1002/gepi.20509](https://doi.org/10.1002/gepi.20509)
- Kruppa J, Ziegler A, König IR (2012) Risk estimation and risk prediction using machine-learning methods. *Hum Genet* 131:1639–1654. doi:[10.1007/s00439-012-1194-y](https://doi.org/10.1007/s00439-012-1194-y)
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037–2048
- Lemeshow S, Hosmer DW (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115:92–106
- Lin X, Song K, Lim N et al (2009) Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score—the CoLaus Study. *Diabetologia* 52:600–608. doi:[10.1007/s00125-008-1254-y](https://doi.org/10.1007/s00125-008-1254-y)
- Lindström S, Schumacher FR, Cox D et al (2012) Common genetic variants in prostate cancer risk prediction—results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol Biomarkers Prev* 21:437–444. doi:[10.1158/1055-9965.EPI-11-1038](https://doi.org/10.1158/1055-9965.EPI-11-1038)
- Lluis-Ganella C, Subirana I, Lucas G et al (2012) Assessment of the value of a genetic risk score in improving the estimation

- of coronary risk. *Atherosclerosis* 222:456–463. doi:[10.1016/j.atherosclerosis.2012.03.024](https://doi.org/10.1016/j.atherosclerosis.2012.03.024)
- Lysenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359:2220–2232. doi:[10.1056/NEJMoa0801869](https://doi.org/10.1056/NEJMoa0801869)
- Mealiffe ME, Stokowski RP, Rhees BK et al (2010) Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* 102:1618–1627. doi:[10.1093/jnci/djq388](https://doi.org/10.1093/jnci/djq388)
- Meigs JB, Shrader P, Sullivan LM et al (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359:2208–2219. doi:[10.1056/NEJMoa0804742](https://doi.org/10.1056/NEJMoa0804742)
- Morrison AC, Bare LA, Chambless LE et al (2007) Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol* 166:28–35. doi:[10.1093/aje/kwm060](https://doi.org/10.1093/aje/kwm060)
- Mühlenbruch K, Jeppesen C, Joost H-G et al (2013) The value of genetic information for diabetes risk prediction—differences according to sex, age, family history and obesity. *PLoS One* 8:e64307. doi:[10.1371/journal.pone.0064307](https://doi.org/10.1371/journal.pone.0064307)
- Nam RK, Zhang WW, Trachtenberg J et al (2009) Utility of incorporating genetic variants for the early detection of prostate cancer. *Clin Cancer Res* 15:1787–1793. doi:[10.1158/1078-0432.ccr-08-1593](https://doi.org/10.1158/1078-0432.ccr-08-1593)
- Paynter NP, Chasman DI, Buring JE et al (2009) Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Ann Intern Med* 150:65–72. doi:[10.7326/0003-4819-150-2-200901200-00003](https://doi.org/10.7326/0003-4819-150-2-200901200-00003)
- Paynter NP, Chasman DI, Paré G et al (2010) Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA* 303:631–637. doi:[10.1001/jama.2010.119](https://doi.org/10.1001/jama.2010.119)
- Pencina MJ, D’Agostino RB, Vasan RS (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. doi:[10.1002/sim.2929](https://doi.org/10.1002/sim.2929) (**discussion 207–12**)
- Pencina MJ, D’Agostino RB, Vasan RS (2010) Statistical methods for assessment of added usefulness of new biomarkers. *Clin Chem Lab Med* 48:1703–1711. doi:[10.1515/CCLM.2010.340](https://doi.org/10.1515/CCLM.2010.340)
- Pepe MS (2011) Problems with risk reclassification methods for evaluating prediction models. *Am J Epidemiol* 173:1327–1335. doi:[10.1093/aje/kwr013](https://doi.org/10.1093/aje/kwr013)
- Pepe M, Janes H (2013) Methods for evaluating prediction performance of biomarkers and tests. In: Lee M-LT, Gail M, Pfeiffer R et al (eds) *Risk Assessment and Evaluation of Predictions SE*, vol 7. Springer New York, pp 107–142
- Peterson TA, Doughty E, Kann MG (2013) Towards precision medicine: advances in computational approaches for the analysis of human variants. *J Mol Biol* 425:4047–4063. doi:[10.1016/j.jmb.2013.08.008](https://doi.org/10.1016/j.jmb.2013.08.008)
- Previtali E, Bucciarelli P, Passamonti SM, Martinelli I (2011) Risk factors for venous and arterial thrombosis. *Blood Transfus* 9:120–138. doi:[10.2450/2010.0066-10](https://doi.org/10.2450/2010.0066-10)
- Rembold CM (1998) Number needed to screen: development of a statistic for disease screening. *BMJ* 317:307–312. doi:[10.1136/bmj.317.7154.307](https://doi.org/10.1136/bmj.317.7154.307)
- Ripatti S, Tikkanen E, Orho-Melander M et al (2010) A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 376:1393–1400. doi:[10.1016/S0140-6736\(10\)61267-6](https://doi.org/10.1016/S0140-6736(10)61267-6)
- Salinas CA, Koopmeiners JS, Kwon EM et al (2009) Clinical utility of five genetic variants for predicting prostate cancer risk and mortality. *Prostate* 69:363–372. doi:[10.1002/pros.20887](https://doi.org/10.1002/pros.20887)
- Schulze MB, Weikert C, Pischon T et al (2009) Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. *Diabetes Care* 32:2116–2119. doi:[10.2337/dc09-0197](https://doi.org/10.2337/dc09-0197)
- Siontis GCM, Tzoulaki I, Siontis KC, Ioannidis JPA (2012) Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* 344:e3318
- Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21:128–138. doi:[10.1097/EDE.0b013e3181c30fb2](https://doi.org/10.1097/EDE.0b013e3181c30fb2)
- Stranger BE, Stahl EA, Raj T (2011) Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187:367–383. doi:[10.1534/genetics.110.120907](https://doi.org/10.1534/genetics.110.120907)
- Tada H, Shiffman D, Smith JG et al (2014) Twelve-single nucleotide polymorphism genetic risk score identifies individuals at increased risk for future atrial fibrillation and stroke. *Stroke* 45:2856–2862. doi:[10.1161/STROKEAHA.114.006072](https://doi.org/10.1161/STROKEAHA.114.006072)
- Talmud PJ, Hingorani AD, Cooper JA et al (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 340:b4838
- Tam CHT, Ho JSK, Wang Y et al (2013) Use of net reclassification improvement (NRI) method confirms the utility of combined genetic risk score to predict type 2 diabetes. *PLoS One*. doi:[10.1371/journal.pone.0083093](https://doi.org/10.1371/journal.pone.0083093)
- Thanassoulis G, Vasan RS (2010) Genetic cardiovascular risk prediction: will we get there? *Circulation* 122:2323–2334. doi:[10.1161/CIRCULATIONAHA.109.909309](https://doi.org/10.1161/CIRCULATIONAHA.109.909309)
- Tikkanen E, Havulinna AS, Palotie A et al (2013) Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. *Arterioscler Thromb Vasc Biol* 33:2261–2266. doi:[10.1161/ATVBAHA.112.301120](https://doi.org/10.1161/ATVBAHA.112.301120)
- Tzoulaki I, Liberopoulos G, Ioannidis JPA (2009) Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 302:2345–2352. doi:[10.1001/jama.2009.1757](https://doi.org/10.1001/jama.2009.1757)
- Vachon CM, Pankratz VS, Scott CG et al (2015) The contributions of breast density and common genetic variation to breast cancer risk. *JNCI J Natl Cancer Inst* 107:dju397. doi:[10.1093/jnci/dju397](https://doi.org/10.1093/jnci/dju397)
- Van Hoek M, Dehghan A, Witteman JCM et al (2008) Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. *Diabetes* 57:3122–3128. doi:[10.2337/db08-0425](https://doi.org/10.2337/db08-0425)
- Vassy JL, Meigs JB (2012) Is genetic testing useful to predict type 2 diabetes? *Best Pract Res Clin Endocrinol Metab* 26:189–201. doi:[10.1016/j.beem.2011.09.002](https://doi.org/10.1016/j.beem.2011.09.002)
- Vassy JL, Dasmahapatra P, Meigs JB et al (2012a) Genotype prediction of adult type 2 diabetes from adolescence in a multi-racial population. *Pediatrics* 130:e1235–e1242. doi:[10.1542/peds.2012-1132](https://doi.org/10.1542/peds.2012-1132)
- Vassy JL, Durant NH, Kabagambe EK et al (2012b) A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia* 55:2604–2612. doi:[10.1007/s00125-012-2637-7](https://doi.org/10.1007/s00125-012-2637-7)
- Vassy JL, Hivert M-F, Porneala B et al (2014) Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes* 63:2172–2182. doi:[10.2337/db13-1663](https://doi.org/10.2337/db13-1663)
- Wacholder S, Hartge P, Prentice R et al (2010) Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 362:986–993. doi:[10.1056/NEJMoa0907727](https://doi.org/10.1056/NEJMoa0907727). **Performance**
- Walford GA, Porneala BC, Dauriz M et al (2014) Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care* 37:2508–2514. doi:[10.2337/dc14-0560](https://doi.org/10.2337/dc14-0560)
- Wang TJ (2011) Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation* 123:551–565. doi:[10.1161/CIRCULATIONAHA.109.912568](https://doi.org/10.1161/CIRCULATIONAHA.109.912568)
- Wang J, Stancáková A, Kuusisto J, Laakso M (2010) Identification of undiagnosed type 2 diabetic individuals by the Finnish Diabetes

- risk score and biochemical and genetic markers: a population-based study of 7232 Finnish men. *J Clin Endocrinol Metab* 95:3858–3862. doi:[10.1210/jc.2010-0012](https://doi.org/10.1210/jc.2010-0012)
- Yusuf S, Hawken S, Ounpuu S et al (2004) Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 364:937–952. doi:[10.1016/S0140-6736\(04\)17018-9](https://doi.org/10.1016/S0140-6736(04)17018-9)
- Zheng SL, Sun J, Wiklund F et al (2008) Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 358:910–919. doi:[10.1056/NEJMoa075819](https://doi.org/10.1056/NEJMoa075819)