Published in final edited form as:
J Med Genet. 2012 September ; 49(9): 601–608. doi:10.1136/jmedgenet-2011-100716.

Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status

Anika Hüsing[1], **Federico Canzian**[2], **Lars Beckmann**[1], **Montserrat Garcia-Closas**[3], **W. Ryan Diver**[4], **Michael J. Thun**[4], **Christine D. Berg**[3], **Robert N. Hoover**[3], **Regina G. Ziegler**[3], **Jonine D. Figueroa**[3], **Claudine Isaacs**[5], **Anja Olsen**[6], **Vivian Viallon**[7], **Heiner Boeing**[8], **Giovanna Masala**[9], **Dimitrios Trichopoulos**[10], **Petra H.M. Peeters**[11], **Eiliv Lund**[12], **Eva Ardanaz**[13], **Kay-Tee Khaw**[14], **Per Lenner**[15], **Laurence N. Kolonel**[16], **Daniel O. Stram**[17], **Loïc Le Marchand**[16], **Catherine A. McCarty**[18], **Julie E. Buring**[19,20], **I-Min Lee**[19], **Shumin Zhang**[20], **Sara Lindström**[21], **Susan E. Hankinson**[20], **Elio Riboli**[22], **David J. Hunter**[21], **Brian E. Henderson**[17], **Stephen J. Chanock**[3], **Christopher A. Haiman**[1], **Peter Kraft**[21], and **Rudolf Kaaks**[1] **on behalf of the BPC3**

[1]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany [2]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany [3]Division of Genetics and Epidemiology, and Breakthrough Breast Cancer Centre, Institute of Cancer Research, London, UK [4]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA [5]Lombardi Comprehensive Cancer Center, Georgetown University, Washington DC, USA [6]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark [7]INSERM U521, Institut Gustave Roussy, Villejuif, France [8]Department of Epidemiology, Deutsches Institut für Ernährungsforschung, Potsdam-Rehbrücke, Germany [9]Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy [10]Department of Epidemiology, Harvard School of Public Health, USA and Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece [11]Julius Center, University Medical Center, Utrecht, The Netherlands [12]Institute of Community Medicine University of Tromsø, Tromsø, Norway [13]Navarre Public Health Institute, Pamplona, Spain and Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), Spain [14]School of Clinical Medicine, University of Cambridge, Cambridge, UK [15]Department of Oncology, Umeå University, Umeå, Sweden [16]Cancer Research Center, University of Hawaii, Honolulu HI, USA [17]University of Southern California, Los Angeles CA, USA [18]The Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI, USA [19]Department of Epidemiology, Harvard School of Public Health, Boston MA, USA [20]Brigham and Women's Hospital and Harvard Medical School, Boston MA, USA [21]Program in Molecular and

Corresponding author: Anika Hüsing, Im Neuenheimer Feld 581, 69120 Heidelberg, Germany, Phone: 49-(0)6221-42-2219, Fax: 49-(0)6221-42-2203, a.huesing@dkfz.de.

Competing interests
No authors have any competing interests to declare.

Author's contributions
AH wrote the statistical analysis plan, cleaned and analysed the data, and drafted and revised the paper. She is guarantor.
FC, LB, MG-C, BEH, SJC, CAH, PK and RK contributed to the design and drafted and revised the paper, MJT, BEH, RGZ, SJC, RK, ER and DJH initiated the cohort consortium project.
WRD, MJT, CDB, RNH, RGZ, JDF, CI, AO, VV, HB, GM, DT, PHMP, EL, EA, K-TK, PL, LNK, DOS, LLeM, CAMcC, JEB, I-ML, SZ, SL, SEH, ER, DJH conducted the epidemiologic studies and contributed samples and covariate data to the BPC3
All authors contributed to the writing of the manuscript.

Genetic Epidemiology, Harvard School of Public Health, Boston MA, USA [22]School of Public Health, Imperial College, London, UK

Abstract

Objective—There is increasing interest in adding common genetic variants identified through genome wide association studies (GWAS) to breast cancer risk prediction models. First results from such models showed modest benefits in terms of risk discrimination. Heterogeneity of breast cancer as defined by hormone-receptor status has not been considered in this context. In this study we investigated the predictive capacity of 32 GWAS-detected common variants for breast cancer risk, alone and in combination with classical risk factors, and for tumors with different hormone receptor status.

Material and Methods—Within the Breast and Prostate Cancer Cohort Consortium (BPC3), we analyzed 6009 invasive breast cancer cases and 7827 matched controls of European ancestry, with data on classical breast cancer risk factors and 32 common gene variants identified through GWAS. Discriminatory ability with respect to breast cancer of specific hormone receptor-status was assessed with the age- and cohort-adjusted concordance statistic ($AUROC_a$). Absolute risk scores were calculated with external reference data. Integrated discrimination improvement (IDI) was used to measure improvements in risk prediction.

Results—We found a small but steady increase in discriminatory ability with increasing numbers of genetic variants included in the model (difference in $AUROC_a$ going from 2.7 to 4%). Discriminatory ability for all models varied strongly by hormone receptor status

Discussion and Conclusion—Adding information on common polymorphisms provides small but statistically significant improvements in the quality of breast cancer risk prediction models. We consistently observed better performance for receptor positive cases, but the gain in discriminatory quality is not sufficient for clinical application.

Keywords

breast cancer; risk prediction; genetic factors; hormone receptor status

OBJECTIVE

Results from genome wide association studies (GWAS) are continuously adding to our knowledge of genetic risk factors for breast cancer [1–13]. Though effects for single gene variants are small, cumulatively they may eventually explain a sizable proportion of heritable breast cancer risk, and there is increasing interest in utilizing information from common genetic polymorphisms for breast cancer risk prediction. Risk prediction models can be an important tool for breast cancer prevention, by identifying women at high risk who would mostly benefit from targeted preventive measures such as mammography screening, or chemoprevention, e.g. with tamoxifen or raloxifene. Present recommendations for identifying women at sufficiently high risk to benefit from chemoprevention include reference to the Breast Cancer Risk Assessment Tool (BCRAT) originally developed by Gail et al. [14] with the aim to reduce costs not only in terms of financial expense, but also to optimize expected medical benefits against possible negative side effects (e.g. increased risk of endometrial cancer) [15]. Likewise, in the light of new results on the limited benefit of mammography screening for some women [16], which needs to be balanced against financial costs as well as possible negative side effects such as radiation and overdiagnosis or false positive diagnosis, it appears worthwhile to also consider the application of risk prediction models in the context of mammography screening [17–19].

The Breast and Prostate Cancer Cohort Consortium (BPC3) offers a large and well characterized study population with both classical epidemiologic risk factor and genetic data [20], which allow the computation and evaluation of comprehensive risk prediction models. Here we present results from this resource, evaluating the collective predictive quality of 32 common gene variants that were reported to be associated with breast cancer in at least one GWAS at genome-wide significance level [1–13]. We investigated risk of breast cancer overall as well as by subtypes defined by estrogen and progesterone receptor status. Besides analyses of the discriminatory potential of genetic and non-genetic risk factor information, we also translated our results to estimates of absolute risk.

MATERIAL AND METHODS

Study population

The BPC3 has been described in detail elsewhere [20]. Briefly, the consortium pools genotyping information and extensive questionnaire data from large, well established prospective cohorts based in the USA and Europe. Cases of invasive primary breast cancer and matched controls were identified from five participating American cohorts: the American Cancer Society Cancer Prevention Study-II (CPS-II) [21], the Harvard Nurses' Health Study (NHS) [22] the Hawaii-Los Angeles Multiethnic Cohort (MEC) [23], the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) [24] and the European Prospective Investigation into Cancer and Nutrition (EPIC) [25]. Depending on the study cohort, cancer cases were identified by linkage with population-based tumor registries and/or self-reported and confirmed through medical records. Controls were matched to cases by ethnicity and age and in some cohorts additional matching criteria were employed, such as current use of HRT at blood donation (EPIC and NHS), recruitment center (EPIC), and further details concerning blood donation (fasting status, time of the day and phase of the menstrual cycle in EPIC, date of blood collection in CPS2 and NHS). Written informed consent was obtained from all subjects, and the project was approved by the appropriate institutional review board for each cohort. In the present analysis, we focused exclusively on subjects of European descent in order to have more homogeneous results, because most breast cancer gene variant discovery has been among women of European ancestry, and because the other ethnic groups represent a comparatively small fraction (18%) of BPC3 subjects.

From 6009 invasive cancers in women of European descent 83% could be classified with respect to estrogen receptor status (21% negative), and 72% could be classified with respect to progesterone receptor (32% negative). Details of measurement and classification of receptor status within the different cohorts are given in the Supplement. To differentiate between breast cancer developed before and after menopause, we regarded cases diagnosed before the age of 55 as predominantly premenopausal, 'early disease onset' (22%), and cases diagnosed after the age of 60 as postmenopausal, 'late disease onset' (62%).

Genetic data

In the current phase of BPC3, 32 single nucleotide polymorphisms (SNPs), that were previously reported as significantly associated with breast cancer risk at a genome-wide significance level ($p<10^{-7}$) were genotyped for a replication study, some results already published [26]. Genotyping assays were designed and performed using Taqman chemistry with reagents by Applied Biosystems (Foster City, CA, USA). Genotyping was performed in four laboratories (located at the University of Southern California, the US National Cancer Institute, Harvard School of Public Health, and the German Cancer Research Center, DKFZ). Laboratory personnel were blinded to case-control status. Within each study, blinded duplicate samples were also included and concordance of these results was greater

than 99%. The genotyping success rate within each cohort was on average 95.6% (range 90.5%–99.4%). For four loci, either the SNP reported in the original study or a surrogate in complete or near complete linkage disequilibrium was genotyped (rs4415084 or surrogate rs920329 (r^2=0.981 in HapMap CEU), rs999737 or surrogate rs10483813 (r^2=1 in HapMap CEU), rs10931936 or surrogate rs700635 (r^2=1 in HapMap CEU), rs1250003 or surrogate rs704010 (r^2=1 in HapMap CEU)). These 32 SNPs include markers that were used for risk prediction in an earlier simulation model by Gail [27], and nine of the ten markers investigated by Wacholder et al [28] based on case and control genotypes in studies that partially overlap the current data set (32% of our subjects). Subjects with a genotyping call rate below 75% were excluded from the genetic data (777 subjects, 5%). Thus our analyses including genetic information were limited to 6.009 cases and 7.827 controls in total. Details on these cases and controls within the different cohorts are given in Table S1 (Supplementary material).

Statistical Methods

Imputation—Classical risk factor information was complete for the majority of subjects (68%), with completeness of single variables ranging from 78% to 99% (data on family history were not available for several study centers in EPIC). To account for the fraction of missing data as summarized in Table S2 (Supplementary material), ten-fold multiple imputation [29]was applied to these covariates, conditioning each covariate on all the others, including case-control status [30]. A maximum of 8 missing genotypes were independently imputed ten times from the SNPs' allele frequency within the cohort (country of origin within EPIC), or overall (rs9383935 was not available for NHS-subjects).

Prediction models—Unconditional logistic regression models were applied throughout this analysis. A subject's most important matching criteria of age-group and cohort membership (country of origin within EPIC) were included in all models to adjust for the matched design.

Application of the traditional model of the BCRAT [14] was not possible with our data, because information on history of former biopsies or benign diseases was not available for all studies in our database. To evaluate the net gain in prediction from the genetic components, we included other well established risk-factors for breast cancer in our study into a model building process to produce an extended covariate model. Thus a sequence of unconditional logistic models with breast cancer status as outcome was fitted to imputed covariate and genotype data:

1. a *covariate* effect model derived through a backwards selection process on available covariates;
2. *genetic* models of seven (as in [27]), nine (as in [28]) and all thirty-two SNPs,or a subgroup of those with independent effect; and
3. combinations of the genetic effects with the covariate model as defined in 1 and 2, to investigate the additional value of the genetic information.

Evaluation of the models

Internal validation—We corrected for overfitting [31, 32] with application of a split-sample design to the multiple imputed data, stratified by cohort (and country within EPIC), using two thirds as training and one third as test data. With age at menarche and age at first full term pregnancy and the adjustment variables age and cohort kept in the model, backwards model selection as provided from SAS (9.2) PROC LOGISTIC was applied to each imputed training data set [33]. The final covariate model included all parameters that

were chosen in more than five of the ten selected models. All model parameters were then again estimated on the training data and without further adjustment applied to the test data.

Evaluation of prediction quality—The statistical fit of models was compared with likelihood-ratio tests and the *Akaike*-criterion within the training data. The Akaike-criterion is adjusting the likelihood of a model for the number of parameters included and thus facilitates the direct comparison of the fit of not necessarily nested models. Discriminative quality of a model to distinguish cases from controls was evaluated with the *AUROC*-statistic derived from predicted values from the test data. Because all data were age-matched with varying case-control ratios in the different cohorts, we calculated the covariate-adjusted $AUROC_a$ [34] from relative risk levels adjusted for age and cohort effects.

Estimation of absolute risks—We used external reference data on age-specific risk rates from cancer registries [35] to transfer our results from models on relative risk to absolute risk levels, representing the probability for a breast-cancer free woman to be diagnosed with breast-cancer in the next five years. Assuming the age-specific covariate distributions within the control subjects to be comparable to the populations covered by these cancer registries, for each demographic group defined by cohort-membership and age-stratum, the absolute baseline risk was calculated from the average relative risk in control subjects. This baseline risk was then applied to all members of this demographic group. Details of these risk calculations are given in the Supplement. Around the threshold advised by the U.S. Preventive Services Task Force [36] to consider chemopreventive treatment from a 5-year risk level of 1.66 %, we built classes of women with risk-score below 1% as low risk, above 1.66% as high risk and above 3.5% as very high risk, to evaluate the potential reclassification gain.

Model comparisons—The change of discrimination of case and non-case subjects due to different risk models was compared stratified by cohort and country within EPIC with the *IDI*, the integrated discrimination improvement, which is independent of class limits [37]. For illustrative purposes we present reclassification tables and the corresponding net reclassification improvement *NRI* [37]. Since older women are over represented in our sample, we also adjusted the age distribution towards that of the US standard population (white only) in 2000 [38].

Stratification over disease subtypes and by age group—To evaluate the disease subtype-specific predictive quality, the prediction models estimated from the full dataset were applied to cases of negative or positive estrogen (ER) or progesterone receptor status (PR) separately with their corresponding matched controls. Because tumor characteristics may differ by the age at which tumors are diagnosed, we also analyzed early and late onset cases, and substrata defined by different ER-status in early and late disease. Our choice of cutpoint was motivated by the concept that tumor development is a relatively long-term process, and tumors diagnosed before the age of 55 would have developed predominantly through a woman's pre-menopausal phase of life.

RESULTS

Evaluation of risk discrimination

Based on the risk factors age at menarche, age at first full term pregnancy, count of full term pregnancies, age at menopause, ever use of hormone replacement therapy, body mass-index in interaction with menopausal status at baseline, smoking and alcohol consumption, our covariate model had a predictive quality in terms of $AUROC_a$ of 56.4% [95% CI: 54.7 –

58.2%]. Definitions and mutually adjusted estimates of model parameters for the classical epidemiological risk factors are given in supplementary Table S3.

With respect to the genetic polymorphisms studied, relative risk estimates generally corresponded to previous findings, except for SNPs rs2180341, rs1011970, rs3817198, rs909116, rs2075555, and rs311499, for which previously observed associations were not replicated in our data (Table 1).

The genetic information of all 32 SNPs combined yielded a discriminative power of $AUROC_a$=58.3% [95% CI: 56.7 – 60.0%]. If only the strongest signal in terms of OR was preserved from SNPs within the same region and six non-significant SNPs were eliminated, this quality was unchanged ($AUROC_a$=58.4%, 95%CI: 56.7 – 60.0%) (Table 2).

The 496 tests of individual pair-wise SNP-interactions resulted in a minimal p-value of 0.0003; thus, within this high-dimensional frame-work we found no evidence to include any genetic interaction terms into the prediction models. Also, combining all genotypes into a simplified genetic score based on the total count of risk-alleles instead of fitting individually weighted SNP effects led to a simplified model, which was inferior to that with individually weighted SNP effects as measured by the *Akaike* information criterion. The best-fitting genetic model was that including information on the 18 SNPs that had statistically significant effects on different loci into a multiple log-additive model with individually weighted per-allele effects for each SNP. Also, comparing between the sets of 7, 9 and 18 SNPs that were included in former studies by Gail [27] and Wacholder et al. [28] and in our present analysis, it can be seen, that each increment in SNP number led to an improvement of the $AUROC_a$, with levels of 56.4, 56.9 and 58.4% respectively.

Adding the 18 SNPs to the covariate model resulted in an $AUROC_a$ of 60.5 [95% CI: 58.9 – 62.2%] and in only a very small improvement in discrimination of 0.16% in terms of IDI.

Although for each of the breast cancer subtypes as defined by ER/PR status there was a statistically significant improvement in model discrimination through addition of genetic model components, this improvement was much smaller for negative receptor tumors than for positive tumors (Table 3).Prediction quality both in terms of $AUROC_a$ and IDI varied more by ER-status than by PR-status..

In addition to the sub-classification by ER/PR status we also found that prediction quality due to genetic factors was generally better for cases with earlier diagnosis, and again this was particularly the case for ER+ tumors (Table S4, supplementary material), where the highest $AUROC_a$ was observed (63.8% [95% CI: 58.8 – 68.9%]).

Estimation of absolute risks and net reclassification improvement

The need to balance risk of different diseases and possible side-effects in the context of preventive actions for breast cancer warrants the generation of absolute risk levels. To simulate the gain from adding genetic information to a model of classical risk factors in practical terms, the estimated absolute risk levels from the covariate model and the model combining genetic and classical risk factors classified according to the cutpoint of 1.66%, as suggested for tamoxifen treatment are given in Table 4.

This shows a significant improvement in terms of NRI of 8.3% (95% CI 5.5–11%), which however would vary slightly when regarding different cut-points around the one presented (data not shown). According to the combined model of covariates plus 18 SNPs, 52% of cases and 40% of controls would then be classified into the “high risk” category (> five-year risk threshold of 1.66%) and might be considered for possible chemoprevention by

tamoxifen treatment according to recent recommendations [15]. This again is related to the age distribution in our sample with a majority of elder women. If the age distribution is weighted corresponding to the US in 2000 [38], the NRI at the 1.66% risk limit is 4.7%.

DISCUSSION

In this analysis of 6.009 invasive breast cancer cases and 7.827 control subjects, all of European ancestry, we found that the genotypes of common SNPs previously shown to be associated with breast cancer risk collectively confer at least as much information for breast cancer risk prediction as an optimized model of classical epidemiologic risk factors ($AUROC_a$ 56.4% vs. 58.3%). Furthermore, in exploiting what is currently the largest prognostic study base with information on genetic and non-genetic (classical) risk factors, we found that adding the genetic information to the classical risk factors leads to a moderate but significant improvement of breast cancer risk overall by adding 3.9% in AUROC. This is similar to findings recently reported from another study [39]. For the most comprehensive risk prediction model, which incorporated the classical risk factor data and 18 significantly associated SNPs, the adjusted concordance statistic $AUROC_a$ was equal to 60.5% for breast cancer overall. To put this into perspective, this improvement in prediction capacity is currently almost as high as what has been estimated for mammographic density [40].

The predictive quality of all models was clearly better for ER-positive diseases, and varied from 55.4% for ER- breast cancer to 61.8% for ER+ cancer. This difference reflects the fact that most of the classical risk factors, as well as the genetic variants identified through breast cancer GWAS studies so far, are predominantly related to risk of hormone-receptor-positive disease [41–45], which constitute the majority of breast cancers in women of European descent. Our findings achieve clinical relevance, because it is known that specific preventive applications may also have different impacts on hormone-receptor-positive and -negative breast cancers. For example in the context of chemoprevention with tamoxifen and raloxifene the incidence of hormone-receptor-positive cancer types is reduced, while the incidence of breast cancer with negative receptor status appears unchanged [15].

Stratifying by age at diagnosis of disease, we saw a slightly better prediction by all genetic models for breast cancer at younger ages, but a lower predictive capacity of the covariates at younger age. The latter reflects that the covariates BMI, HRT-use and age at menopause have an effect only after menopause. Again, this age dependence of the predictive quality of our models was mostly present for ER+ breast cancer.

The genetic information provided best discrimination of risk when the effects of the SNPs were weighted according to specific risk estimates instead of being grouped into a common allele-counting score; this finding is in line with recent findings by Hsu et al. [46]. Although single SNP effects were small, a steady increase in prediction quality was observed with growing number of genetic markers included in the prediction model. This indicates potential for further improvement in prediction models as more genetic markers will be identified through GWAS studies.

Though there was a statistically significant improvement in discrimination quality, this may still be too small to be considered meaningful for clinical application, weighted against current genotyping costs and additional education, which would be needed to prepare both physicians and patients for the consideration of genotyping results. Proper evaluation of this aspect would require a cost-benefit analysis as done by Gail [18] and Freedman [47], based on explicit assumptions for benefits and costs related to positive and negative classifications. Although such a balance is beyond the scope of this work, we did, for illustrative purposes, derive absolute risk levels, that generally form the basis of such cost benefit analyses, and

calculated the theoretical gain from the adding genotyping information in terms of a net reclassification improvement (NRI). As shown by our example, even the small increase in discrimination quality of our model due to the genotype data could lead to a net reclassification improvement of 8.3% overall. This estimate, however, while illustrative, must be interpreted with caution, as the NRI generally depends strongly on absolute risk cut points used.

Strengths of our study are its prospective design and its overall study size, both characteristics that are desirable for developing statistical models on the joint effects of classical epidemiologic and genetic risk factors. Also, as the cohorts represent population groups from both North America and Europe, we could estimate models spanning heterogeneous risk factor distributions over a relatively large range, especially for the classical risk factors. An important observation, in this context, is that the genetic factors showed stable effects on breast cancer risk across all study groups [26].

As a few risk factors considered standard part of the BCRAT-model were not available from some of the participating studies in BPC3, namely family history, the number of past breast biopsies and history of benign breast disease, we could not fit this predefined model. However, the question that we intended to address was, whether genetic factors could predict risk independently and additionally to traditional risk factors.

We also had no information on mammographic breast density, which is another important predictor for breast cancer risk [40, 48]. We thus could not examine whether the improvement of risk prediction by including information from the genetic polymorphisms would be the same in the presence of these additional risk factors. For the subset of subjects who provided information on family history we compared the predictive capacity of our models and we saw no evidence for a difference between subjects who confirmed a family history of breast cancer and those without.

A further limitation of our study, which resulted from its nested case-control design, is that within each of the contributing cohorts control subjects had been matched to the cases by age. Thus, we could only estimate $AUROC_a$ after age adjustment, and the discriminatory effect of age itself on breast cancer risk prediction could not be estimated. As a consequence, our results in terms of absolute $AUROC$ values were lower in comparison to other studies where age was included as a predictive variable, although it can be argued that since increasing age is a risk factor for most cancers, age itself is not a useful discriminator of risk between two people of the same age. In our estimations of absolute risks, however, differences due to age and population effects were entered back in on the basis of age-specific risk levels from different, regional cancer registries.

Finally, a word of caution is needed with regard to our model estimates for absolute breast cancer risk. We presented our absolute risk models for a mixed North-American/European population, which is a theoretical construct, and did not correct for competing risks from other sources. Also, while we used cancer registry data that were specific for European and US sub-populations included into the contributing cohorts of BPC3, the validity of absolute risk estimates depends on the assumption that risk factor distributions within the cohorts were identical to those in the general populations covered by the cancer registries. If, in reality, the distributions of genetic and other risk factors were different, absolute risk estimates may be improperly calibrated. Moreover, the thresholds we regarded for model evaluation in terms of reclassification and NRI are not generally applicable because the decision to use tamoxifen also depends on risks of non-cancer complications, such as stroke and pulmonary embolism, and consequently higher thresholds may be more appropriate for

older women. Thus while quite instructive in showing possible gain in classification accuracy for “real-life” purposes, these results have to be interpreted with great caution.

Conclusion

In conclusion, our analyses indicate, that small increases in prediction quality may be expected with a growing number of genetic markers detected to be associated with breast cancer risk. For the gene variants identified so far, these increases in predictive quality can be observed particularly for ER-positive breast cancer subtypes. The quality of risk prediction overall of genetic and classical risk factors combined is still far from a level to allow accurate discrimination of prospective cases or non-cases for preventive measures on a population level. However, extrapolating from our observations and considering further theoretical estimations [49] it can be anticipated that the discriminative power will further increase as the number of known common genetic determinants of breast cancer grows.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by US National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233-07 to D.J.H.; U01-CA98710-06 to M.J.T.; U01-CA98216-06 to E.R. and R.K.; and U01-CA98758-07 to B.E.H.) and Intramural Research Program of National Institutes of Health and National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Reference List

1. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tomczyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos SS I, Lathrop M, Houlston RS, Peto J. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst. 2011 Mar 2; 103(5):425–435. [PubMed: 21263130]
2. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PD, Stratton MR, Dunning AM, Rahman N, Easton DF. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010 Jun; 42(6):504–507. [PubMed: 20453838]
3. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le ML, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, Mannermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007 Jun 28; 447(7148): 1087–1093. [PubMed: 17529967]

4. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr. Hoover RN, Thomas G, Chanock SJ. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007 Jul; 39(7):870–874. [PubMed: 17529973]

5. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le ML, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, Frigge ML, Geller F, Gudbjartsson D, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Jonsson T, von HS, Werelius B, Margolin S, Lindblom A, Mayordomo JI, Haiman CA, Kiemeney LA, Johannsson OT, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007 Jul; 39(7):865–869. [PubMed: 17529974]

6. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MR, Giles GG, Fletcher O, Johnson N, dos SS I, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, Risch A, Wang-Gohrke S, Schurmann P, Bogdanova N, Dork T, Fagerholm R, Aaltonen K, Blomqvist C, Nevanlinna H, Seal S, Renwick A, Stratton MR, Rahman N, Sangrajrang S, Hughes D, Odefrey F, Brennan P, Spurdle AB, Chenevix-Trench G, Beesley J, Mannermaa A, Hartikainen J, Kataja V, Kosma VM, Couch FJ, Olson JE, Goode EL, Broeks A, Schmidt MK, Hogervorst FB, Van't Veer LJ, Kang D, Yoo KY, Noh DY, Ahn SH, Wedren S, Hall P, Low YL, Liu J, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Sigurdson AJ, Stredrick DL, Alexander BH, Struewing JP, Pharoah PD, Easton DF. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet. 2007 Mar; 39(3):352–358. [PubMed: 17293864]

7. Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, Antman K, Hwang SJ. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Med Genet. 2007; 8(Suppl 1):S6. [PubMed: 17903305]

8. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, Jakobsdottir M, Bergthorsson JT, Gudmundsson J, Aben KK, Strobbe LJ, Swinkels DW, van Engelenburg KC, Henderson BE, Kolonel LN, Le ML, Millastre E, Andres R, Saez B, Lambea J, Godino J, Polo E, Tres A, Picelli S, Rantala J, Margolin S, Jonsson T, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Sveinsdottir SG, Alexiusdottir K, Saemundsdottir J, Sigurdsson A, Kostic J, Gudmundsson L, Kristjansson K, Masson G, Fackenthal JD, Adebamowo C, Ogundiran T, Olopade OI, Haiman CA, Lindblom A, Mayordomo JI, Kiemeney LA, Gulcher JR, Rafnar T, Thorsteinsdottir U, Johannsson OT, Kong A, Stefansson K. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2008 Jun; 40(6):703–706. [PubMed: 18438407]

9. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J, Offit K. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A. 2008 Mar 18; 105(11):4340–4345. [PubMed: 18326623]

10. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee IM, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr. Hoover RN, Chanock SJ, Hunter DJ. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009 May; 41(5):579–584. [PubMed: 19330030]

11. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W, Shu XO. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009 Mar; 41(3):324–328. [PubMed: 19219042]
12. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos SS I, Peto J, Stratton MR, Rahman N, Jacobs K, Prentice R, Anderson GL, Rajkovic A, Curb JD, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver WR, Bojesen S, Nordestgaard BG, Flyger H, Dork T, Schurmann P, Hillemanns P, Karstens JH, Bogdanova NV, Antonenkova NN, Zalutsky IV, Bermisheva M, Fedorova S, Khusnutdinova E, Kang D, Yoo KY, Noh DY, Ahn SH, Devilee P, van Asperen CJ, Tollenaar RA, Seynaeve C, Garcia-Closas M, Lissowska J, Brinton L, Peplonska B, Nevanlinna H, Heikkinen T, Aittomaki K, Blomqvist C, Hopper JL, Southey MC, Smith L, Spurdle AB, Schmidt MK, Broeks A, van Hien RR, Cornelissen S, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Schmutzler RK, Burwinkel B, Bartram CR, Meindl A, Brauch H, Justenhoven C, Hamann U, Chang-Claude J, Hein R, Wang-Gohrke S, Lindblom A, Margolin S, Mannermaa A, Kosma VM, Kataja V, Olson JE, Wang X, Fredericksen Z, Giles GG, Severi G, Baglietto L, English DR, Hankinson SE, Cox DG, Kraft P, Vatten LJ, Hveem K, Kumle M, Sigurdson A, Doody M, Bhatti P, Alexander BH, Hooning MJ, van den Ouweland AM, Oldenburg RA, Schutte M, Hall P, Czene K, Liu J, Li Y, Cox A, Elliott G, Brock I, Reed MW, Shen CY, Yu JC, Hsu GC, Chen ST, Anton-Culver H, Ziogas A, Andrulis IL, Knight JA, Beesley J, Goode EL, Couch F, Chenevix-Trench G, Hoover RN, Ponder BA, Hunter DJ, Pharoah PD, Dunning AM, Chanock SJ, Easton DF. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009 May; 41(5):585–590. [PubMed: 19330027]
13. Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, Segre AV, McGee K, McGuffog L, Kartsonaki C, Morrison J, Healey S, Sinilnikova OM, Stoppa-Lyonnet D, Mazoyer S, Gauthier-Villars M, Sobol H, Longy M, Frenay M, GEMO SC, Hogervorst FB, Rookus MA, Collee JM, Hoogerbrugge N, van Roozendaal KE, Piedmonte M, Rubinstein W, Nerenstone S, Van LL, Blank SV, Caldes T, de la Hoya M, Nevanlinna H, Aittomaki K, Lazaro C, Blanco I, Arason A, Johannsson OT, Barkardottir RB, Devilee P, Olopade OI, Neuhausen SL, Wang X, Fredericksen ZS, Peterlongo P, Manoukian S, Barile M, Viel A, Radice P, Phelan CM, Narod S, Rennert G, Lejbkowicz F, Flugelman A, Andrulis IL, Glendon G, Ozcelik H, Toland AE, Montagna M, D'Andrea E, Friedman E, Laitman Y, Borg A, Beattie M, Ramus SJ, Domchek SM, Nathanson KL, Rebbeck T, Spurdle AB, Chen X, Holland H, John EM, Hopper JL, Buys SS, Daly MB, Southey MC, Terry MB, Tung N, Overeem Hansen TV, Nielsen FC, Greene MH, Mai PL, Osorio A, Duran M, Andres R, Benitez J, Weitzel JN, Garber J, Hamann U, Peock S, Cook M, Oliver C, Frost D, Platte R, Evans DG, Lalloo F, Eeles R, Izatt L, Walker L, Eason J, Barwell J, Godwin AK, Schmutzler RK, Wappenschmidt B, Engert S, Arnold N, Gadzicki D, Dean M, Gold B, Klein RJ, Couch FJ, Chenevix-Trench G, Easton DF, Daly MJ, Antoniou AC, Altshuler DM, Offit K. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. PLoS Genet. 2010 Oct.6(10):e1001183. [PubMed: 21060860]
14. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989 Dec 20; 81(24):1879–1886. [PubMed: 2593165]
15. Visvanathan K, Chlebowski RT, Hurley P, Col NF, Ropka M, Collyar D, Morrow M, Runowicz C, Pritchard KI, Hagerty K, Arun B, Garber J, Vogel VG, Wade JL, Brown P, Cuzick J, Kramer BS, Lippman SM. American society of clinical oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. J Clin Oncol. 2009 Jul 1; 27(19):3235–3258. [PubMed: 19470930]
16. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2011 Jan. %19;1:CD001877.:CD001877.
17. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008 Jun 26; 358(26):2796–2803. [PubMed: 18579814]
18. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. J Natl Cancer Inst. 2009 Jul 1; 101(13):959–963. [PubMed: 19535781]

19. Pashayan N, Duffy SW, Chowdhury S, Dent T, Burton H, Neal DE, Easton DF, Eeles R, Pharoah P. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. Br J Cancer. 2011 May 10; 104(10):1656–1663. [PubMed: 21468051]
20. Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, Haynes RB, Henderson BE, Kaaks R, Stram DO, Thomas G, Thun MJ, Blanche H, Buring JE, Burtt NP, Calle EE, Cann H, Canzian F, Chen YC, Colditz GA, Cox DG, Dunning AM, Feigelson HS, Freedman ML, Gaziano JM, Giovannucci E, Hankinson SE, Hirschhorn JN, Hoover RN, Key T, Kolonel LN, Kraft P, Le ML, Liu S, Ma J, Melnick S, Pharaoh P, Pike MC, Rodriguez C, Setiawan VW, Stampfer MJ, Trapido E, Travis R, Virtamo J, Wacholder S, Willett WC. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer. 2005 Dec; 5(12):977–985. [PubMed: 16341085]
21. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, Feigelson HS, Thun MJ. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer. 2002 May 1; 94(9):2490–2501. [PubMed: 12015775]
22. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. Nat Rev Cancer. 2005 May; 5(5):388–396. [PubMed: 15864280]
23. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 2000 Feb 15; 151(4):346–357. [PubMed: 10695593]
24. Hayes RB, Reding D, Kopp W, Subar AF, Bhat N, Rothman N, Caporaso N, Ziegler RG, Johnson CC, Weissfeld JL, Hoover RN, Hartge P, Palace C, Gohagan JK. Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. Control Clin Trials. 2000 Dec; 21(6 Suppl):55S–57S.
25. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-De-Mesquita HB, Peeters PH, Lund E, Engeset D, Gonzalez CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002 Dec; 5(6B):1113–1124. [PubMed: 12639222]
26. Campa D, Kaaks R, Le ML, Haiman CA, Travis RC, Berg CD, Buring JE, Chanock SJ, Diver WR, Dostal L, Fournier A, Hankinson SE, Henderson BE, Hoover RN, Isaacs C, Johansson M, Kolonel LN, Kraft P, Lee IM, McCarty CA, Overvad K, Panico S, Peeters PH, Riboli E, Sanchez MJ, Schumacher FR, Skeie G, Stram DO, Thun MJ, Trichopoulos D, Zhang S, Ziegler RG, Hunter DJ, Lindstrom S, Canzian F. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. J Natl Cancer Inst. 2011 Aug 17; 103(16):1252–1263. [PubMed: 21791674]
27. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J Natl Cancer Inst. 2008 Jul 16; 100(14):1037–1041. [PubMed: 18612136]
28. 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 Mar 18; 362(11):986–993. [PubMed: 20237344]
29. Schafer JL. Multiple imputation: a primer 1. Stat Methods Med Res. 1999 Mar; 8(1):3–15. [PubMed: 10347857]
30. Raghunathan, Trivellore E.; Lepkowski, James M.; Van Hoewykand, John; Solenberger, Peter. A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Survey Methodology. 2010; 27(1):85–95.
31. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. Breast Cancer Res Treat. 1998; 52(1–3):289–303. [PubMed: 10066088]
32. Harrell FE Jr. Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996 Feb 28; 15(4):361–387. [PubMed: 8668867]

33. Steyerberg EW, Harrell FE Jr. Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001 Aug; 54(8):774–781. [PubMed: 11470385]
34. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. Am J Epidemiol. 2008 Jul 1; 168(1):89–97. [PubMed: 18477651]
35. Curado, MP.; Edwards, B.; Shin, HR.; Ferlay, JHM.; Boyle, P. Cancer Incidence in Five Continents. Vol. 9. Lyon, IARC: IARC Scientific Publications; 2007. Ref Type: Generic
36. U.S.Preventive Services Task Force. Chemoprevention of breast cancer: recommendations and rationale. U.S.Preventive Services Task Force. 2002; 137(1) Ref Type: Online Source.
37. Pencina M, D'Agostino R, D'Agostino R, Vasan R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in medicine. 2008 Jan 30; 27(2):157–172. [PubMed: 17569110]
38. Day, JC. Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050. Washington, DC: U.S. Bureau of the Census,U.S. Government Printing Office; 1996. p. P25-P1130.Report No
39. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. J Natl Cancer Inst 2010. 3 Nov; 102(21):1618–1627.
40. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008 Mar 4; 148(5):337–347. [PubMed: 18316752]
41. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Zamora P, Brauch H, Justenhoven C, Hamann U, Ko YD, Bruening T, Haas S, Dork T, Schurmann P, Hillemanns P, Bogdanova N, Bremer M, Karstens JH, Fagerholm R, Aaltonen K, Aittomaki K, von SK, Blomqvist C, Mannermaa A, Uusitupa M, Eskelinen M, Tengstrom M, Kosma VM, Kataja V, Chenevix-Trench G, Spurdle AB, Beesley J, Chen X, Devilee P, van Asperen CJ, Jacobi CE, Tollenaar RA, Huijts PE, Klijn JG, Chang-Claude J, Kropp S, Slanger T, Flesch-Janys D, Mutschelknauss E, Salazar R, Wang-Gohrke S, Couch F, Goode EL, Olson JE, Vachon C, Fredericksen ZS, Giles GG, Baglietto L, Severi G, Hopper JL, English DR, Southey MC, Haiman CA, Henderson BE, Kolonel LN, Le ML, Stram DO, Hunter DJ, Hankinson SE, Cox DG, Tamimi R, Kraft P, Sherman ME, Chanock SJ, Lissowska J, Brinton LA, Peplonska B, Klijn JG, Hooning MJ, Meijers-Heijboer H, Collee JM, van den Ouweland A, Uitterlinden AG, Liu J, Lin LY, Yuqing L, Humphreys K, Czene K, Cox A, Balasubramanian SP, Cross SS, Reed MW, Blows F, Driver K, Dunning A, Tyrer J, Ponder BA, Sangrajrang S, Brennan P, McKay J, Odefrey F, Gabrieau V, Sigurdson A, Doody M, Struewing JP, Alexander B, Easton DF, Pharoah PD. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 2008 Apr.4(4):e1000054. [PubMed: 18437204]
42. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev. 2004 Oct; 13(10):1558–1568. [PubMed: 15466970]
43. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis. Int J Cancer. 2009 Feb 1; 124(3): 698–712. [PubMed: 18988226]
44. Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD. Common genetic variation in candidate genes and susceptibility to subtypes of breast cancer. Cancer Epidemiol Biomarkers Prev. 2009 Jan; 18(1):255–259. [PubMed: 19124506]
45. Reeves GK, Travis RC, Green J, Bull D, Tipper S, Baker K, Beral V, Peto R, Bell J, Zelenika D, Lathrop M. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. JAMA. 2010 Jul 28; 304(4):426–434. [PubMed: 20664043]
46. Hsu FC, Sun J, Zhu Y, Kim ST, Jin T, Zhang Z, Wiklund F, Kader AK, Zheng SL, Isaacs W, Gronberg H, Xu J. Comparison of two methods for estimating absolute risk of prostate cancer

based on single nucleotide polymorphisms and family history. Cancer Epidemiol Biomarkers Prev. 2010 Apr; 19(4):1083–1088. [PubMed: 20332264]

47. Freedman AN, Yu B, Gail MH, Costantino JP, Graubard BI, Vogel VG, Anderson GL, McCaskill-Stevens W. Benefit/risk assessment for breast cancer chemoprevention with raloxifene or tamoxifen for women age 50 years or older. J Clin Oncol. 2011 Jun 10; 29(17):2327–2333. [PubMed: 21537036]
48. Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst. 2006 Sep 6; 98(17):1215–1226. [PubMed: 16954474]
49. Gu W, Pepe MS. Estimating the capacity for improvement in risk prediction with a marker. Biostatistics. 2009 Jan; 10(1):172–186. [PubMed: 18714084]

Table 1

Results from univariate replication analysis on full dataset: SNPs given with frequency of risk allele (RAF), genotyping call rate (cr), per-allele odds ratio (OR) for breast cancer with 95% confidence interval (95%CI), for breast cancer overall and for ER- and ER+ tumors alone.

						All		ER +		ER −	
Gene	**Rs-number**	**Chr**	**Position**	**RAF**	**cr**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
NOTCH2	**RS11249433**[s]	1	120982136	0.42	98.3	1.09	(1.04–1.15)	1.13	(1.07–20)	1.00	(0.89–1.12)
CASP8	RS10931936[**]	2	201852173	0.29	97.0	1.08	(1.03–1.13)	1.04	(0.98–11)	1.10	(0.97–1.25)
CASP8	**RS1045485**[s]	2	201857834	0.87	98.1	1.13	(1.05–1.21)	1.13	(1.03–23)	1.12	(0.95–1.32)
Intergenic	**RS13387042**[s]	2	217614077	0.53	98.2	1.20	(1.15–1.26)	1.26	(1.19–34)	1.05	(0.94–1.17)
SLC4A7	**RS4973768**[s]	3	27391017	0.49	98.7	1.08	(1.04–1.14)	1.09	(1.03–15)	0.97	(0.87–1.08)
TERT	**RS10069690**[s]	5	1279790	0.26	97.0	1.04	(0.99–1.09)	1.04	(0.97–11)	1.18	(1.05–1.34)
Intergenic	RS4415084[**]	5	44698272	0.41	97.9	1.08	(1.03–1.13)	1.11	(1.04–17)	1.02	(0.91–1.14)
Intergenic	**RS10941679**[s]	5	44742255	0.26	97.3	1.12	(1.07–1.18)	1.16	(1.08–23)	1.03	(0.91–1.17)
MAP3K1	**RS889312**[s]	5	56067641	0.29	98.4	1.11	(1.06–1.17)	1.14	(1.07–22)	1.02	(0.90–1.15)
ECHDC1 / RNF146	RS2180341	6	127642323	0.76	98.1	1.03	(0.98–1.09)	1.02	(0.96–10)	1.01	(0.89–1.15)
C6orf97 (ESR1)	RS9383935	6	151939848	0.09	57.4	1.11	(1.01–1.22)	1.12	(0.99–26)	0.96	(0.73–1.26)
Intergenic	RS3757318	6	151955806	0.08	98.2	1.13	(1.04–1.22)	1.16	(1.04–28)	1.07	(0.87–1.33)
Intergenic	RS9383938	6	151987357	0.09	99.0	1.12	(1.04–1.21)	1.15	(1.05–27)	1.12	(0.92–1.37)
Intergenic	**RS2046210**[s4]	6	151990059	0.35	98.4	1.09	(1.04–1.14)	1.10	(1.04–17)	1.07	(0.96–1.20)
Intergenic	RS13281615	8	128424801	0.42	98.1	1.08	(1.03–1.14)	1.09	(1.02–15)	1.08	(0.96–1.20)
Intergenic	**RS1562430**[s]	8	128457034	0.59	98.9	1.12	(1.07–1.17)	1.16	(1.09–22)	1.13	(1.01–1.27)
CDKN 2BAS	RS1011970	9	22052134	0.17	98.5	1.07	(1.01–1.14)	1.00	(0.93–08)	1.02	(0.88–1.18)
Intergenic	**RS865686**[s]	9	110888478	0.64	98.6	1.10	(1.05–1.15)	1.10	(1.04–16)	1.20	(1.07–1.35)
Intergenic	**RS2380205**[s]	10	5926740	0.56	98.8	1.05	(1.01–1.10)	1.04	(0.98–10)	1.04	(0.92–1.16)

						All		ER +		ER −	
Gene	**Rs-number**	**Chr**	**Position**	**RAF**	**cr**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
ZNF365	**RS10995190**^{s}	10	63948688	0.86	98.2	1.12	(1.05–1.19)	1.12	(1.04–21)	1.06	(0.90–1.26)
ZNF365	RS16917302	10	64261198	0.91	97.5	1.04	(0.97–1.12)	1.06	(0.97–16)	0.97	(0.80–1.18)
ZMIZ1	**RS1250003**^{**s}	10	80846814	0.39	97.0	1.04	(1.00–1.09)	1.05	(0.99–11)	0.98	(0.87–1.10)
FGFR2	RS3750817	10	123322567	0.61	97.7	1.16	(1.11–.22)	1.18	(1.11–26)	1.03	(0.92–1.15)
FGFR2	**RS2981582**^{s}	10	123342308	0.41	98.1	1.22	(1.17–1.28)	1.24	(1.17–31)	1.09	(0.97–1.22)
LSP1	RS3817198	11	1865583	0.68	97.7	1.01	(0.97–1.07)	1.01	(0.95–07)	1.07	(0.95–1.20)
LSP1	RS909116	11	1898522	0.53	98.8	1.04	(1.00–1.09)	1.05	(0.99–10)	0.99	(0.89–1.11)
Intergenic	**RS614367**^{s}	11	69037945	0.15	99.2	1.15	(1.08–1.22)	1.18	(1.10–27)	0.98	(0.83–1.15)
RAD51L1	**RS999737**^{**s}	14	68104435	0.77	98.7	1.12	(1.06–1.18)	1.14	(1.07–22)	1.05	(0.92–1.20)
TNRC9	**RS3803662**^{s}	16	51143843	0.29	98.2	1.20	(1.14–1.26)	1.18	(1.11–26)	1.19	(1.05–1.34)
COL1A1	RS2075555	17	45629290	0.14	98.8	1.04	(0.97–1.11)	1.05	(0.97–14)	1.14	(0.97–1.34)
COX11	**RS6504950**^{s}	17	50411470	0.73	98.9	1.08	(1.02–1.14)	1.10	(1.03–17)	1.08	(0.96–1.23)
GMEB2	RS311499	20	62217589	0.93	97.2	1.03	(0.94–1.12)	1.05	(0.94–17)	0.96	(0.76–1.20)

^{s}these 18 SNPs were selected as independent effects into common model

^{s4}SNP selected on basis of common model on 4 SNPs in that region

**for some cohorts genotypes not from this SNP but from surrogate marker with $r^2 \geq 0.98$.

Table 2

Discriminative value $AUROC_a$ (95%-confidence interval) for models including different covariates and genetic effects, and integrated discrimination improvement (IDI) due to addition of genetic effects to the covariate model.

Cov-effect	None		Covariates[c]	
geneffect	AUROC (95% CI)	IDI	AUROC (95% CI)	IDI
None	0.5 *		0.564 (0.547 – 0.581)	
32 SNPs	0.583 (0.567 – 0.600)	0.16%	0.604 (0.588 – 0.621)	0.17%
18 SNPs **	0.584 (0.567 – 0.600)	0.15%	0.605 (0.589 – 0.622)	0.16%
9 SNPs	0.569 (0.552 – 0.586)	0.11%	0.595 (0.579 – 0.612)	0.12%
7 SNPs	0.564 (0.547 – 0.581)	0.10%	0.591 (0.574 – 0.608)	0.10%

* by construction

** better than 32 SNPs according to Akaike information criterion

[c] including parameters on age at menarche, at first birth and at menopause and count of births, BMI, alcohol consumption, smoking and use of hormone replacement therapy.

Table 3

Discrimination quality $AUROC_a$ (with 95% confidence interval) and integrated discrimination improvement (*IDI*), after addition of 32 or 18 SNPs to the null model and the covariate model, in different disease strata.

Genetic Effect	ER+	ER−	PR+	PR-	Early diagnosis	Late diagnosis
# cases in full data	3920	1059	2953	1381	1316	3747
No covariates						
+ 32 SNPs	0.596 (0.574 – 618)	0.530 (0.493 – 568)	0.598 (0.573 – 0.622)	0.560 (0.527 – 594)	0.609 (0.573 – 0.645)	0.574 (0.553 – 596)
IDI	***0.208%***	***0.047%***	***0.208%***	***0.101%***	***0.115%***	***0.191%***
+18 SNPs	0.595 (0.574 – 617)	0.530 (0.492 – 567)	0.597 (0.573 – 0.621)	0.560 (0.526 – 593)	0.610 (0.574 – 0.645)	0.574 (0.552 – 596)
IDI	***0.199%***	***0.050%***	***0.197%***	***0.095%***	***0.114%***	***0.192%***
Covariate model[c]	0.570 (0.547 – 592)	0.544 (0.507 – 581)	0.570 (0.545 – 0.595)	0.544 (0.510 – 579)	0.540 (0.502 – 0.577)	0.562 (0.539 – 584)
+ 32 SNPs	0.618 (0.596 – 639)	0.553 (0.516 – 590)	0.619 (0.595 – 0.643)	0.580 (0.547 – 614)	0.615 (0.579 – 0.651)	0.594 (0.572 – 616)
IDI	***0.215%***	***0.051%***	***0.216%***	***0.108%***	***0.125%***	***0.202%***
+18 SNPs	0.618 (0.596 – 639)	0.554 (0.517 – 591)	0.619 (0.595 – 0.643)	0.582 (0.549 – 615)	0.615 (0.579 – 0.651)	0.595 (0.573 – 616)
IDI	***0.204%***	***0.054%***	***0.207%***	***0.099%***	***0.120%***	***0.202%***

[c] including parameters on age at menarche, at first birth and at menopause and count of births, BMI, alcohol consumption, smoking and use of hormone replacement therapy.

J Med Genet. Author manuscript; available in PMC 2013 October 09.

Table 4

Reclassification in our test sample into absolute risk classes due to the effect of adding 18 significant SNPs to covariate model (NRI=8.3%), correctly reclassified cases and controls across the 5-year-risk threshold of 1.66% are indicated as bold (NRI across all 4 categories:15.8%).

		Risk score from covariates alone			
Risk score including 18 SNPs		**Risk<1%**	**Risk >1%, <1.66%**	**1.66% <= Risk < 3.5%**	**Risk > 3.5%**
Risk<1%	% cases	10.77	6.55	0.69	0
	% controls	15.61	9.19	**1.18**	0
Risk >1%, <1.66%	% cases	3.33	16.28	9.93	0.05
	% controls	2.28	17.55	**13.75**	0
1.66% <= Risk < 3.5%	% cases	**0.4**	**11.17**	32.26	0.5
	%controls	0.23	7.37	28.75	0.23
Risk > 3.5%	% cases	0	**0.15**	7.54	0.4
	% controls	0	0.11	3.46	0.3