Published in final edited form as:

Epidemiology. 2012 July ; 23(4): 616–622. doi:10.1097/EDE.0b013e31825583a0.

Assessing Disease Risk in Genome-wide Association Studies Using Family History

Arpita Ghosh, Patricia Hartge, Mark P. Purdue, Stephen J. Chanock, Laufey Amundadottir, Zhaoming Wang, Nicolas Wentzensen, Nilanjan Chatterjee, and Sholom Wacholder
Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, USA

Abstract

We show how to use reports of cancer in family members to discover additional genetic associations or confirm previous findings in genome-wide association (GWA) studies conducted in case-control, cohort, or cross-sectional studies. Our novel family-history-based approach allows economical association studies for multiple cancers, without genotyping of relatives (as required in family studies), follow-up of participants (as required in cohort studies), or oversampling of specific cancer cases, (as required in case-control studies). We empirically evaluate the performance of the proposed family-history-based approach in studying associations with prostate and ovarian cancers, using data from GWA studies previously conducted within the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. The family-history-based method may be particularly useful for investigating genetic susceptibility to rare diseases, for which accruing cases may be very difficult, by using disease information from non-genotyped relatives of participants in multiple case-control and cohort studies designed primarily for other purposes.

In many epidemiologic studies exploring genetic association, participants provide information at enrollment on history of cancers in family members, particularly first-degree relatives. Investigators commonly use this family history information to control for confounding in standard logistic regression. Thornton et al.¹ have suggested using family history for optimally weighting cases and controls in association studies as a way to improve power. Our novel approach for using family history in genome-wide association (GWA) studies exploits Mendelian rules of inheritance to detect genetic associations. In essence, we infer the relatives' genotype distributions from the observed genotypes of participants, using patterns of Mendelian inheritance. The test for association between relatives' inferred genotypes and their disease status is, in fact, exactly the same as the test for association between participants' genotypes and relatives' disease status. Hence, in practice, we compare the genotype distribution of the participants with and without family history to provide evidence of association with the underlying risk of disease.

We illustrate the method using as examples studies of prostate and ovarian cancer in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. We test for associations

Corresponding author's name and mailing address, telephone and fax numbers, and e-mail address: Arpita Ghosh, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd, Room 5054, Rockville, MD 20852, Tel: (301) 443-8228, ghosha3@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

with prostate cancer reported in fathers, brothers, or sons of men and women in the cohort, and associations with ovarian cancer reported in mothers, sisters or daughters of the same cohort members. For prostate cancer we compare results with the standard GWA case-control results for participants in the Cancer Screening trial. Because some participants were genotyped for a nested case-control study of prostate cancer and most were genotyped for unrelated studies, we were able to compare the results from a classic case-control design to the family-history design.

Methods

Study Population

Details of the design and other features of the Cancer Screening trial have been described previously.² Briefly, this is an ongoing randomized clinical trial to test the effect of screening on mortality for prostate, lung, colorectal and ovarian cancers. 154,901 men and women aged 55–74 years were enrolled between 1993 and 2001 and randomized to either the screening or the control arm.

Family history variable

In two self-administered questionnaires (one at baseline and one about ten years after enrolment began), participants were asked about history of cancer in first-degree relatives and half-siblings. In the baseline questionnaire (<https://www.plcostars.com/Public/Documents/PLCO/BQF.pdf>), respondents were asked to list family members (including parents, full and half-siblings, and children) with a history of cancer, and to indicate the cancer type and age at diagnosis. Of the 154,901 trial participants, 149,980 (97%) completed the baseline questionnaire. The supplemental questionnaire (<https://www.plcostars.com/Public/Documents/PLCO/SQX.pdf>), completed by 104,007 (69%) of 149,980 participants in 2006–2007, collected information on history of breast, prostate, lung, ovarian, lymphoma, colorectal, endometrial, bladder, leukemia, and “other cancers” in first-degree relatives.

We defined the family-history variable for prostate cancer as positive if one or more male first-degree relatives were reported to have prostate cancer. We considered family history to be positive if the respondent indicated family history of prostate cancer in either the baseline or supplemental questionnaires, and negative if both responses were “no” or “missing.” We defined the family-history variable for ovarian cancer correspondingly.

We pooled information over the two questionnaires to estimate the number of male first-degree relatives for the participants and the number of those reported to have prostate cancer. We did not have complete family structure from the baseline questionnaire; the questions did not ask about the number of sons for each participant, nor did they differentiate between full and half-brothers. The supplemental questionnaire asked for the number of full brothers and sons of each participant. From this we were able to calculate the average number of full brothers and sons for each participant, and apply that average to estimate the number of full brothers and sons for all participants. In counting the number of relatives reported to have prostate cancer, we considered a male relative to be affected if he was reported in either of the two questionnaires. We followed the same rules for ovarian cancer.

Genotyping

Blood was collected for the screening arm participants, and buccal cell samples were collected from control arm participants. Using these samples, 6,695 participants (4%) were genotyped as part of several previously published and ongoing GWA studies (prostate

cancer component of the Cancer Genetic Markers of Susceptibility project,^{3–5} the Pancreatic Cancer Cohort Consortium,^{6,7} and a multi-stage GWA study of bladder cancer^{8,9}). We used a list of single-nucleotide polymorphisms (SNPs) from <http://www.genome.gov/gwastudies/> (accessed July 20, 2010) that have been reported to be associated with prostate cancer. We restricted the analysis to a single SNP per locus when two or more SNPs within a locus were in strong linkage disequilibrium. For ovarian cancer, we examined a list of 44 previously published and subsequently confirmed SNPs.^{10–12} Genotypes for 11 of the 44 ovarian SNPs were imputed using the program IMPUTE v2.¹³ We used 1000 Genomes project (June 2010 release) and HapMap 3 (release 2), both based on the human genome build 36, as the imputation reference set.

Statistical Analysis

We restricted our analyses to white participants ($n=6,411$; 96% of all participants with GWA data). For the family-history-based method, we fit a logistic regression model for each SNP, with the SNP genotype as an independent variable and the family history of the participant as the dependent variable. The test for association from the family-history-based model corresponds to an indirect test for association between the SNP and the underlying risk of prostate or ovarian cancer, with a participant's family as the unit of observation. For comparison with the disease-based GWA approach, we performed the conventional direct test for association for the prostate-related SNPs with the disease status of the participant as the response. We considered the 1151 participants with prostate cancer (identified as part of Cancer Genetic Markers of Susceptibility Prostate Cancer GWA study as described elsewhere³), as cases and the remaining 5260 genotyped participants as controls for studying disease-based association. Under the null hypothesis of no association between the SNP genotype and the disease, the family-history-based test maintains the correct Type I error, as does the standard case-control test for association.

For both models, we assumed the additive mode of inheritance. The logistic regression on imputed SNPs used the expected count instead of the exact count of minor alleles.¹³ We included the top five principal components (based on a selected panel of about 12,000 independent SNPs¹⁴) in the regression to control for population stratification.¹⁵ For the family-history-based regression model we adjusted for the participants': sex and age at enrollment, treating age as a continuous covariate. For the disease-based model we adjusted for age at enrollment and family history of prostate cancer. All statistical analyses were performed using SAS, version 9.1 (SAS Institute, Cary, NC).

Results

Characteristics of the participants are reported in Table 1. Of the 6411 participants available for analyses, 738 (12%) participants reported having one or more first-degree relatives with prostate cancer. The participants included 2243 control subjects and 4168 cases from one or more GWA studies. The first column of p-values in Table 1 next to each cancer type corresponds to a test of whether cases for that cancer type are more likely to have reported history of prostate cancer in their families than the rest. As expected, the 1151 men selected as cases for prostate cancer GWA study reported substantially more family history of prostate cancer than the remaining 5260 participants (19% vs. 10%; one-sided p-value = 1.1×10^{-16}), while the participants chosen as cases for other cancer types were not more likely to have a family history of prostate cancer.

Three hundred and two (5%) participants out of 6411 reported one or more first-degree relatives with ovarian cancer. Study participants chosen as cases for GWA studies of bladder and kidney cancers were more likely to have family history of ovarian cancer than the remaining participants (5.5% in 1682 bladder cancer cases vs. 4.4% in remaining 4729

participants, one-sided p-value = 0.03; 7.5% in 281 kidney cancer cases vs. 4.6% in remaining 6130 participants, one-sided p-value = 0.01).

Three-fourths (77%) of participants were men. A greater percentage of men than women reported family history of prostate cancer (12% vs. 10%; two-sided p-value = 0.03). However, when we excluded the 1151 men included in the case series of the prostate cancer GWA study (23% among 4955 men in our analysis), there was no longer a difference in reported family history by sex (two-sided p-value = 0.97). Participants with and without a history of prostate cancer among first-degree relatives had similar age distributions (mean age = 63.8 vs. 64.1; two-sided p-value = 0.12).

Women were more likely than men to report one or more family members with ovarian cancer (7.4% vs. 3.9% respectively; two-sided p-value = 2.9×10^{-8}), as was also noted previously by Pinsky et al.¹⁶ Participants with and without ovarian-cancer-affected relatives had similar age distributions (mean age = 64.4 vs. 64.1; two-sided p-value = 0.22).

The history of prostate and ovarian cancer among the first-degree relatives of the 6411 participants is generally consistent with what is known about the epidemiology of these cancers. The 4,010 participants who responded to the supplemental questionnaire reported averages of 1.5 brothers, 1.4 sisters, 1.2 sons and 1.2 daughters. Participants reporting family history of prostate cancer had more brothers and sons on average than the participants reporting no family history (2.0 vs. 1.4 brothers; 1.3 vs. 1.2 sons). Similarly, participants with reports of ovarian cancer in family members had more sisters and daughters than participants without (2.1 vs. 1.4 sisters; 1.5 vs. 1.2 daughters). Parents were most likely to be reported as having had prostate (7% of fathers) or ovarian cancer (6% of mothers), with a slightly lower percentage of siblings (4% of brothers; 3% of sisters) and a negligible percentage of sons (0.1%) and daughters (0.3%). We also examined these counts separately for the 1,151 participants selected as cases for prostate cancer GWA study and the remaining 5260 participants, and observed similar patterns. As expected, the prostate cancer Study cases were more likely to report a family history of prostate cancer (6% of first-degree relatives had prostate cancer) than the remaining 5260 participants (3%) or all 6411 participants (3%).

Table 2 presents the p-values for associations with each of the selected 32 prostate cancer SNPs, and the corresponding odds ratio estimates, for both the family-history-based and disease-based methods. The family-history-based p-values are below 0.05 for 6 of these established prostate cancer SNPs and marginally above 0.05 for 3 additional SNPs. Specifically, for the most statistically significant SNP (rs4242382) the family-history-based p-value is 7.4×10^{-5} ; for the second and third most significant SNPs (rs1859962 and rs1512268) the p-values are 0.002 and 0.004 respectively. The disease-based method yielded p-values smaller than 0.05 for many of the previously confirmed SNPs. Comparing the two sets of p-values, those based on family history were weaker (p-value for one-sided Wilcoxon signed-rank test comparing matched pairs of p-values = 0.001).

For three SNPs (rs12621278, rs1512268 and rs620861), which were confirmed with subsequent replication,^{4,17,18} the family-history-based method provided evidence of association when the disease-based method was not informative. For these three SNPs, the family-history-based p-values were 0.02, 3.5×10^{-3} , and 0.04; while the disease-based p-values were 0.47, 0.16 and 0.18.

The odds ratio estimates from the family-history-based method are biased estimates of effect sizes and should be interpreted with caution. A scatter plot of the log odds ratio estimates from the two methods (eFigure 1, <http://links.lww.com>) shows that the estimates go mostly in the same direction, as one would expect. For SNPs with estimates in the same direction,

there is a tendency for the family history-based estimates to be smaller in magnitude than the disease-based estimates.

We also evaluated the family-history-based method for studying association with ovarian cancer. Table 3 presents the p-values for testing associations with the 44 ovarian SNPs via the family-history-based method. None of our participants were genotyped for a nested case-control study of ovarian cancer, and so we could not make a standard disease-based comparison as for prostate cancer. Eleven of the 12 SNPs at 9p22, identified by Song et al.¹² as susceptibility markers for ovarian cancer, were associated with the risk of disease at 0.05 level of significance. Two SNPs (rs2301301 at 2q31 and rs17138237 at 7p21) out of 30 in 9 loci examined by Goode et al.,¹⁰ yielded p-values less than 0.05.

Discussion

Our work shows that information routinely collected about family history in GWA studies can provide new and independent evidence of associations between genetic variants and disease at low marginal cost. The most exciting application of this family-history-based method is the feasibility of a GWA study of a rare disease using genotypes from persons previously studied in case-control or cohort studies for other purposes, when history of the rare disease in family members is available. The method can also add to the information available from cohort members or cases and controls; the method allows one to learn about genetic associations with a sex-specific disease, such as ovarian cancer, using male as well as female study participants.

Like the kin-cohort method,^{19,20} the family-history-based method indirectly assesses genetic susceptibility using error-in-variable methods facilitated by knowledge of the error mechanism from Mendelian principles.²¹ The kin-cohort method estimates penetrance, or risk of disease, in those with a high-risk genotype, whereas the proposed family-history-based approach tests for genetic association. Although the odds ratio estimate from the family-history-based approach is a biased estimate of the effect of the SNP on disease risk, the method provides an unbiased test for association.

Evidence from the family-history-based method may be useful either for discovery of associations or as supplementary information when prior data already exist. The information provided by the family-history-based approach can add power to conventional studies or to pooled analyses. When used by itself, the usual cautions for GWA studies apply.

Thornton et al.¹ used family history information, together with the disease status of the study participants, to gain power in a standard analysis of case-control data. These authors exploited the information on disease status in relatives to construct optimal weights for study participants. Our family-history based approach uses family history of relatives, notably older ones (for adult cancer), for studying association. This approach does not use disease status of the participants, which may be unavailable at the start of a prospective cohort study. We emphasize our ability to study diseases that are not the focus of the main study; a case-control study of one disease can contribute information about risks of many others, and a cohort study that enrolls middle-aged subjects can be useful for investigating associations with a late-onset disease such as prostate cancer or Alzheimer disease, without having to wait for cases to accrue in the cohort.

When comparing the p-values from the family-history-based and disease-based analyses, we should note that the prostate-cancer cases were part of the discovery set for the prostate cancer GWA study. This naturally biases the disease-based analysis favorably for the SNPs that were discovered by that study and gives a spurious power advantage over the family-based analysis. However, there is real attenuation in power for the family-history-based

method due to measurement error from using participants' genotypes in studies of relatives' disease, and this attenuation is far more important. For example, one might genotype men and collect information on their fathers' disease status. The weak correlation between their genotypes contributes to the loss in power. Other sources of power loss are inaccurate reports of disease status of family members leading to phenotype misclassification, and the smaller number of men with affected fathers than in case-control studies, where cases are oversampled. This attenuation in power may be partially countered by larger sample sizes or by collecting information from several family members.

We used simulations to compare the performances of the family-history-based and disease-based methods. For odds ratios ranging from 1.2 to 1.8, we studied relative efficiencies (defined as the inverse ratio of sample sizes needed for 70% power at 0.05 level of significance) of family-history-based analyses, as we increase the number of family members, compared with the standard case-control analysis. We fixed disease prevalence at 8% and considered a SNP with 30% minor allele frequency for our simulations. The family-history-based test with a single first-degree relative requires approximately 4 times the sample size as a standard case-control test for association; as we gather information on more relatives, this ratio decreases (eFigure 2, <http://links.lww.com>). To check the sensitivity of our observations across various parameters, we performed these simulations for different values of disease prevalence and minor allele frequency (eTable 1, <http://links.lww.com>). We found that disease prevalence is the primary determinant of the comparative performance of the family-history-based and case-control analysis.

The quality of the reports on relatives' disease history is the central concern in using reported family history. In the absence of electronically linkable medical records or on-line family history tools, accuracy of study participants' reports of relatives' disease history will vary by characteristics of the participants, the relatives and the disease. Mai et al.²² found that reports on breast, prostate, colorectal and lung cancer in family members of participants in the population-based 2001 Connecticut Family Health Study had low-to-moderate sensitivity and positive predictive value, but high specificity and negative predictive value. Participants' knowledge of disease in family members varied with the disease (reported history of breast cancer in family members had the highest sensitivity, while colorectal cancer had the lowest) and with the degree of relatedness to the participant (reports on first-degree relatives were more accurate than reports on second-degree relatives). Accuracy of reports also depends on other factors: disease status, age, ethnicity, sex and family size of the participant, as well as age and sex of the family members. Family history was not directly verified in the Cancer Screening Trial, but Pinsky et al.¹⁶ indirectly assessed validity of reported family history by comparing reported rates of various cancers in family members with expected rates derived from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database. Overall, the authors observed a ratio of reported rates to expected rates of approximately 0.7; for most cancers, this ratio was in the range of 0.6–1.0 for women and 0.3–0.8 for men. Incomplete or inaccurate family history information collected from participants can reduce power and, when reporting accuracy is differential between study participants with and without family history (as is most likely the case), can induce false-positive reports.

Different family-history variables might be appropriate for different settings. We show that the family-history-based method works with even the simplest yes/no family history variable, which might be the most common in existing studies and is the cheapest to obtain for future questionnaires. We observe that participants reporting family history of cancer had more first-degree relatives on average than participants with no family history; this suggests that formulations of family history taking into account family size (such as proportion of affected family members) may perform better. With additional details, we can construct a

family history score²³ based on family structure and risk covariates for family members. As a first step, we fitted a polytomous logistic regression (eTable 2, column 3 [http://links.lww.com]) treating the response variable (family history of prostate cancer) as a nominal variable (none, one, and multiple family members with prostate cancer); we saw no substantial improvement.

We assume an additive mode of inheritance in our analyses. To check sensitivity of our results to the modeling assumptions, we also fit the general two-degrees-of-freedom model (eTable 2, columns 4 and 5 [http://links.lww.com]). The observed pattern holds even under the general model. The two-degrees-of-freedom model changes the p-values only marginally in an analysis that does not adjust for population substructure (eTable 2, columns 6 and 7 [http://links.lww.com]).

Information on family history of diseases can add value to association studies without additional genotyping. Family history information can supplement disease-based associations, and can be particularly useful when deciding on a set of variants for follow-up in further studies. It may be possible to increase the power of conventional disease-based association studies by combining information on disease status for participants and their relatives, accounting for correlation. In this context, the quasi-likelihood score test proposed by Thornton et al.¹ can be used to construct pedigree-based weights and thus improve power. The power gain from combining family history with disease status may be particularly useful in assessing SNPs with borderline significance based on standard analysis of the study population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Leveraging family history for GWAS

- **Sources of financial support, including granting agency and grant number if applicable:** NA
- **Acknowledgements (including colleagues who contributed to the research but who do not meet the requirements for authorship):** NA

References

1. Thornton T, McPeck MS. Case-Control Association Testing with Related Individuals: A More Powerful Quasi-Likelihood Score Test. *American journal of human genetics*. 2007; 81(2):321–337. [PubMed: 17668381]
2. Philip CP, Gerald LA, Robert SB, Sandra SB, David C, Crawford ED, Ronald F. Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled clinical trials*. 2000; 21(6):273S–309S. [PubMed: 11189684]
3. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007; 39(5): 645–649. [PubMed: 17401363]
4. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nature Genetics*. 2009; 41(10):1055–1057. [PubMed: 19767755]
5. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K. Multiple loci identified in a genome-wide association study of prostate cancer. *Nature genetics*. 2008; 40(3):310–315. [PubMed: 18264096]

6. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, Bueno-de-Mesquita HB. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet.* 2009; 41(9):986–990. [PubMed: 19648918]
7. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, Arslan AA. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22. 1, 1q32. 1 and 5p15. 33. *Nature genetics.* 2010; 42(3):224–228. [PubMed: 20101243]
8. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, Real FX. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet.* 2010; 42(11):978–984. [PubMed: 20972438]
9. Purdue MP, Johansson M, Zelenika D, Toro JR, Scelo G, Moore LE, Prokhorchouk E. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13. 3. *Nature Genetics.* 2010
10. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, Widschwendter M, Vierkant RA, Larson MC, Kjaer SK, Birrer MJ, Berchuck A, Schildkraut J, Tomlinson I, Kiemeny LA, Cook LS, Gronwald J, Garcia-Closas M, Gore ME, Campbell I, Whittemore AS, Sutphen R, Phelan C, Anton-Culver H, Pearce CL, Lambrechts D, Rossing MA, Chang-Claude J, Moysich KB, Goodman MT, Dork T, Nevanlinna H, Ness RB, Rafnar T, Hogdall C, Hogdall E, Fridley BL, Cunningham JM, Sieh W, McGuire V, Godwin AK, Cramer DW, Hernandez D, Levine D, Lu K, Iversen ES, Palmieri RT, Houlston R, van Altena AM, Aben KKH, Massuger LFAG, Brooks-Wilson A, Kelemen LE, Le ND, Jakubowska A, Lubinski J, Medrek K, Stafford A, Easton DF, Tyrer J, Bolton KL, Harrington P, Eccles D, Chen A, Molina AN, Davila BN, Arango H, Tsai Y-Y, Chen Z, Risch HA, McLaughlin J, Narod SA, Ziogas A, Brewster W, Gentry-Maharaj A, Menon U, Wu AH, Stram DO, Pike MC, Beesley J, Webb PM. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet.* 2010; 42(10):874–879. [PubMed: 20852632]
11. Bolton KL, Tyrer J, Song H, Ramus SJ, Notaridou M, Jones C, Sher T, Gentry-Maharaj A, Wozniak E, Tsai Y-Y, Weidhaas J, Paik D, Van Den Berg DJ, Stram DO, Pearce CL, Wu AH, Brewster W, Anton-Culver H, Ziogas A, Narod SA, Levine DA, Kaye SB, Brown R, Paul J, Flanagan J, Sieh W, McGuire V, Whittemore AS, Campbell I, Gore ME, Lissowska J, Yang HP, Medrek K, Gronwald J, Lubinski J, Jakubowska A, Le ND, Cook LS, Kelemen LE, Brook-Wilson A, Massuger LFAG, Kiemeny LA, Aben KKH, van Altena AM, Houlston R, Tomlinson I, Palmieri RT, Moorman PG, Schildkraut J, Iversen ES, Phelan C, Vierkant RA, Cunningham JM, Goode EL, Fridley BL, Kruger-Kjaer S, Blaaker J, Hogdall E, Hogdall C, Gross J, Karlan BY, Ness RB, Edwards RP, Odunsi K, Moysich KB, Baker JA, Modugno F, Heikkinen T, Butzow R, Nevanlinna H, Leminen A, Bogdanova N, Antonenkova N, Doerk T, Hillemanns P, Durst M, Runnebaum I, Thompson PJ, Carney ME, Goodman MT, Lurie G, Wang-Gohrke S, Hein R, Chang-Claude J, Rossing MA, Cushing-Haugen KL, Doherty J, Chen C, Rafnar T, Besenbacher S, Sulem P, Stefansson K, Birrer MJ, Terry KL, Hernandez D, Cramer DW, Vergote I, Amant F, Lambrechts D, Despierre E, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet.* 2010; 42(10):880–884. [PubMed: 20852633]
12. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R, Dork T, Goode EL, Goodman MT, Schildkraut JM, Sellers T, Baglietto L, Beckmann MW, Beesley J, Blaakaer J, Carney ME, Chanock S, Chen Z, Cunningham JM, Dicks E, Doherty JA, Durst M, Ekici AB, Fenstermacher D, Fridley BL, Giles G, Gore ME, De Vivo I, Hillemanns P, Hogdall C, Hogdall E, Iversen ES, Jacobs IJ, Jakubowska A, Li D, Lissowska J, Lubinski J, Lurie G, McGuire V, McLaughlin J, Medrek K, Moorman PG, Moysich K, Narod S, Phelan C, Pye C, Risch H, Runnebaum IB, Severi G, Southey M, Stram DO, Thiel FC, Terry KL, Tsai Y-Y, Tworoger SS, Van Den Berg DJ, Vierkant RA, Wang-Gohrke S, Webb PM, Wilkens LR, Wu AH, Yang H, Brewster W, Ziogas A, Houlston R, Tomlinson I, Whittemore AS, Rossing MA, Ponder BAJ, Pearce CL, Ness RB, Menon U, Kjaer SK, Gronwald J, Garcia-Closas M, Fasching PA, Easton DF, Chenevix-Trench G, Berchuck A, Pharoah PDP, Gayther SA. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet.* 2009; 41(9):996–1000. [PubMed: 19648919]
13. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010; 11(7):499–511. [PubMed: 20517342]

14. Yu K, Wang Z, Li Q, Wacholder S, Hunter D, Hoover R, Chanock S, Thomas G. Population substructure and control selection in genome-wide association studies. *PLoS One*. 2008; 3(7): 2551.
15. Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38(8):904–909. [PubMed: 16862161]
16. Pinsky P, Kramer B, Reding D, Buys S. Reported family history of cancer in the prostate, lung, colorectal, and ovarian cancer screening trial. *American Journal of Epidemiology*. 2003; 157(9): 792. [PubMed: 12727673]
17. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nature Genetics*. 2009; 41(10):1116–1121. [PubMed: 19767753]
18. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S. Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nature Genetics*. 2009; 41(10):1058–1060. [PubMed: 19767752]
19. Wacholder S, Hartge P, Struewing J, Pee D, McAdams M, Brody L, Tucker M. The kin-cohort study for estimating penetrance. *American Journal of Epidemiology*. 1998; 148(7):623. [PubMed: 9778168]
20. Struewing J, Hartge P, Wacholder S, Baker S, Berlin M, McAdams M, Timmerman M, Brody L, Tucker M. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *New England Journal of Medicine*. 1997; 336(20):1401. [PubMed: 9145676]
21. Elston, RC.; Olson, JM.; Palmer, L. *Biostatistical genetics and genetic epidemiology*. John Wiley & Sons Inc; 2002.
22. Mai PL, Garceau AO, Graubard BI, Dunn M, McNeel TS, Gonsalves L, Gail MH, Greene MH, Willis GB, Wideroff L. Confirmation of family cancer history reported in a population-based survey. *Journal of the National Cancer Institute*. 2011; 103(10):788. [PubMed: 21562245]
23. Khoury, M.; Beaty, T.; Cohen, B. *Fundamentals of genetic epidemiology*. USA: Oxford University Press; 1993.

TABLE 1
Characteristics of the Participants by Family History of Prostate and Ovarian Cancer

Characteristics		Family History of Prostate Cancer			Family History of Ovarian Cancer			Total		
		No. Family members ^a with Prostate Cancer		P-values	No. Family members ^b with Ovarian Cancer		P-values			
		0	1+ ^c		0	1+ ^c				
		(n=5673)	(n=738)		(n=6109)	(n=302)		(n=6411)		
Reason for Genotyping	GWA Controls	2029	214	9.5	2146	97	4.32	2243	35	
		Prostate Cancer	938	213	18.5	1100	51	4.43	1151	18
		Bladder Cancer	1501	181	10.8	1589	93	5.53	1682	26.2
		Breast Cancer	5	0	0	4	1	20.00	5	0.1
		Glioma	108	13	10.7	116	5	4.13	121	1.9
	Participants chosen as GWA cases of ^d	Kidney Cancer	249	32	11.4	260	21	7.47	281	4.4
		Lung Cancer	722	69	8.7	758	33	4.17	791	12.3
		Pancreatic Cancer	168	21	11.1	185	4	2.12	189	2.9
		Women	1311	145	10	1348	108	7.42	1456	22.7
		Men	4362	593	12	4761	194	3.92	4955	77.3
Age at enrollment (years); mean (SD)		64.1 (5.12)	63.8 (5.28)		0.12 ^h	64.1 (5.14)	64.4 (5.12)	0.22 ^h	64.1 (5.14)	

^aFather, brothers, and sons

^bMother, sisters, and daughters

^cNo. (%), except where indicated

^dSome participants (n = 51) were chosen as cases for multiple cancer types

^eOne-sided p-value for score test

^fOne-sided p-value for Fisher's exact test

^gTwo-sided p-value for Pearson's chi-squared test

^hTwo-sided p-value for two-sample t-test

TABLE 2
Association of the 32 Prostate Cancer SNPs in 26 Chromosomal Regions with the Risk for Prostate Cancer

SNP	Region	Reported Gene(s)	Minor Allele	Minor Allele Frequency (%)	P-values		Odds Ratio	
					Family-history-based	Disease-based	Family-history-based	Disease-based
rs721048	2p15	EHBP1	A	19	0.644	NA	0.76	NA
rs1465618	2p21	THADA	A	21	0.350	0.871	1.06	1.01
rs12621278	2q31.1	ITGA6	G	6	0.023	0.466	0.74	0.93
rs17181170	3p12.1	NR	A	49	0.070	0.015	0.90	0.89
rs2660753	3p12.1	Intergenic	T	12	0.114	0.217	1.14	1.09
rs10934853 ^a	3q21.3	Intergenic	A	28	0.142	0.004	1.09	1.16
rs12500426	4q22.3	PDLIM5	A	46	0.983	0.131	1.00	1.07
rs7679673	4q24	TET2	A	41	0.510	0.003	1.05	0.86
rs10498792	6p12.2	PKHD1	C	11	0.908	0.860	0.99	0.99
rs9364554	6q25.3	SLC22A3	T	29	0.610	0.396	0.97	1.04
rs10486567	7p15.2	JAZF1	A	23	0.686	0.007	1.03	0.86
rs6465657	7q21.3	JAZF1	C	46	0.526	0.257	0.97	1.05
rs1512268	8p21.2	NKX3.1	A	43	0.004	0.158	1.18	1.07
rs1016343	8q24.21	Intergenic	T	21	0.205	0.005	1.09	1.17
rs4242382	8q24.21	Intergenic	A	11	7.4E-05	9.1E-07	1.38	1.40
rs6983267 ^a	8q24.21	Intergenic	T	49	0.159	7.2E-06	0.92	0.81
rs620861 ^a	8q24	-	T	37	0.014	0.173	0.87	0.94
rs1571801	9q33.2	-	A	26	0.798	0.027	1.02	1.12
rs10993994 ^a	10q11.23	MSMB	T	39	0.156	0.013	1.08	1.12
rs3123078	10q11.23	NR	C	44	0.529	0.023	1.04	1.11
rs4962416	10q26.13	CTBP2	C	29	0.711	4.3E-05	1.02	1.23
rs7127900	11p15.5	IGF2, IGF2AS, INS, TH	A	19	0.381	0.075	1.06	1.11
rs10896449 ^a	11q13.2	Intergenic	A	49	0.061	0.001	0.90	0.85
rs11228565 ^a	11q13.2	Intergenic	A	19	0.892	NA	1.07	NA

SNP	Region	Reported Gene(s)	Minor Allele	Minor Allele Frequency (%)	P-values		Odds Ratio	
					Family-history-based	Disease-based	Family-history-based	Disease-based
rs7931342 ^a	11q13.2	Intergenic	T	49	0.058	0.001	0.90	0.86
rs4054823	17p12	-	C	45	0.847	0.120	1.01	0.93
rs4430796	17q12	Intergenic	G	47	0.253	0.268	0.94	0.95
rs1859962 ^a	17q24.3	NR	G	49	0.002	0.001	1.19	1.17
rs2735839	19q13.33	KLK3	A	15	0.271	0.049	1.09	0.88
rs9623117	22q13.1	TNRC6B	C	23	0.509	0.048	0.96	1.12
rs5759167	22q13.2	NR	T	50	0.023	0.001	0.88	0.86
rs5945572 ^a	Xp11.22	NUDT10, NUDT11, LOC340602, GSPT2, MAGED1	A	37	0.722	0.001	0.98	1.12
Gray cells have p-values less than 0.05								

^a SNPs discovered by the Cancer Genetic Markers of Susceptibility Prostate Cancer GWA study

TABLE 3

Association of Ovarian Cancer SNPs with the Risk for Ovarian Cancer

Source	SNP	Region	Minor Allele	Minor Allele Frequency (%)	Family-history-based P-values
Song et al. ¹²	rs3814113	9p22	C	32	0.019
	rs4445329 ^a	9p22	A	29	0.012
	rs10810666	9p22	T	20	0.007
	rs10962656	9p22	A	15	0.007
	rs12379183	9p22	G	22	0.002
	rs2153271	9p22	C	42	0.020
	rs7861573	9p22	A	22	0.002
	rs10756819	9p22	G	34	0.001
	rs1416742	9p22	G	38	0.583
	rs12379687	9p22	T	16	0.005
	rs4961501	9p22	T	25	0.007
	rs1339552	9p22	T	42	0.032
Bolton et al. ¹¹	rs8170	19p13	A	19	0.391
	rs2363956	19p13	T	49	0.102
Goode et al. ¹⁰	rs2072590	2q31	A	33	0.210
	rs2665390	3q25	C	8	0.589
	rs10088218	8q24	A	13	0.539
	rs9303542 ^a	17q21	G	24	0.583
	rs7517829	1p36	G	41	0.573
	rs2744721	1p36	T	39	0.659
	rs7521902	1p36	A	23	0.837
	rs3920498 ^a	1p36	C	16	0.722
	rs12730089 ^a	1p31	G	31	0.307
	rs7512893	1p31	T	13	0.620
	rs12131772 ^a	1p31	G	12	0.637

Source	SNP	Region	Minor Allele	Minor Allele Frequency (%)	Family-history-based P-values
	rs4670981	2p22	T	13	0.940
	rs12469949	2p22	A	12	0.916
	rs10181603	2p22	T	10	0.964
	rs17024636 ^a	2p22	T	9	0.974
	rs2301301	2q31	C	31	0.041
	rs7793050	7p21	T	23	0.124
	rs10249817	7p21	T	48	0.153
	rs17138237	7p21	A	37	0.039
	rs1516982	8q24	G	15	0.917
	rs10098821	8q24	T	11	0.595
	rs12791486	11p14	G	11	0.431
	rs12800697	11p14	A	11	0.909
	rs12794435 ^a	11p14	A	10	0.844
	rs17241923	11p14	A	11	0.842
	rs17680229 ^a	17q21	A	8	0.451
	rs2084881	17q21	A	28	0.548
	rs7219985	17q21	T	28	0.551
	rs1533057	17q21	A	22	0.912
	rs8072770	17q21	T	22	0.907
Gray cells have p-values less than 0.05					

^aImputed SNPs: 1000 Genomes project (June 2010 release) and HapMap 3 (release 2), both based on the human genome build 36, were used as the imputation reference set.