A Genome-Wide Association Study Identifies A New Ovarian Cancer Susceptibility Locus On 9p22.2

Honglin Song[1,**], Susan J. Ramus[2,**], Jonathan Tyrer[1], Kelly L. Bolton[1,17], Aleksandra Gentry-Maharaj[2], Eva Wozniak[2], Hoda Anton-Culver[3], Jenny Chang-Claude[4], Daniel W. Cramer[5], Richard DiCioccio[6], Thilo Dörk[7], Ellen L. Goode[8], Marc T Goodman[9], Joellen M Schildkraut[10], Thomas Sellers[11], Laura Baglietto[12,13], Matthias W. Beckmann[14], Jonathan Beesley[15], Jan Blaakaer[16], Michael E Carney[9], Stephen Chanock[17], Zhihua Chen[11], Julie M. Cunningham[8], Ed Dicks[1], Jennifer A. Doherty[18], Matthias Dürst[19], Arif B. Ekici[20], David Fenstermacher[11], Brooke L. Fridley[8], Graham Giles[12,13], Martin E. Gore[21], Immaculata De Vivo[22], Peter Hillemanns[7], Claus Hogdall[23], Estrid Hogdall[24], Edwin S Iversen[25], Ian J Jacobs[2], Anna Jakubowska[26], Dong Li[3], Jolanta Lissowska[27], Jan Lubiński[26], Galina Lurie[9], Valerie McGuire[28], John McLaughlin[29], Krzysztof Mędrek[26], Patricia G. Moorman[10], Kirsten Moysich[30], Steven Narod[31], Catherine Phelan[11], Carole Pye[1], Harvey Risch[32], Ingo B Runnebaum[19], Gianluca Severi[12,13], Melissa Southey[33], Daniel O. Stram[34], Falk C. Thiel[14], Kathryn L. Terry[5], Ya-Yu Tsai[11], Shelley S. Tworoger[22], David J. Van Den Berg[34], Robert A. Vierkant[8], Shan Wang-Gohrke[35], Penelope M. Webb[15], Lynne R. Wilkens[9], Anna H Wu[34], Hannah Yang[17], Wendy Brewster[36], Argyrios Ziogas[3], Australian Cancer (Ovarian) Study[37], The Australian Ovarian Cancer Study Group[37], The Ovarian Cancer Association Consortium, Richard Houlston[38], Ian Tomlinson[39], Alice S Whittemore[28], Mary Anne Rossing[18], Bruce A.J. Ponder[1], Celeste Leigh Pearce[34], Roberta B. Ness[40], Usha Menon[2], Susanne Krüger Kjaer[24], Jacek Gronwald[26], Montserrat Garcia-Closas[17], Peter A. Fasching[14,41], Douglas F Easton[42], Georgia Chenevix-Trench[15], Andrew Berchuck[10], Paul D.P. Pharoah[1,*], and Simon A. Gayther[2]

[1] CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK [2] Department of Gynaecological Oncology, UCL EGA Institute for Women's Health, University College London, UK [3] Department of Epidemiology, University of California Irvine, USA [4] Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany [5] Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Boston, MA [6] Department of Cancer Genetics and the Gilda Radner Familial Ovarian Cancer Registry Buffalo, NY, USA [7] Clinics of Obstetrics and Gynaecology, Hannover Medical School, Hannover, Germany [8] Mayo Clinic College of Medicine, Rochester, MN, USA [9] Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, USA [10] The Comprehensive Cancer Center, Duke University Medical Center, Durham, NC, USA [11] Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL, USA [12] Cancer Epidemiology Centre, The Cancer Council Victoria, Rathdowne Street, Carlton VIC 3053, Australia [13] Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Swanston Street, Carlton VIC 3053, Australia [14] University Hospital

*To whom correspondence should be addressed: Dr. Paul Pharoah, CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK., paul1@srl.cam.ac.uk, Tel: 0044 1223 740166.
**Contributed equally to the study.

Erlangen, Department of Gynecology and Obstetrics, Germany [15] The Queensland Institute of Medical Research, Post Office Royal Brisbane Hospital, Australia [16] Department of Gynaecological and Obstetrics, Skejby University Hospital, Arhus [17] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA [18] Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [19] Department of Gynaecology and Obstetrics, University Hospital of Jena, Germany [20] Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany [21] The Royal Marsden Hospital, Gynecological Oncology Unit, London, UK [22] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA [23] The Gyneacologic Clinic, The Juliane Marie Centre, Rigshospitalet, Copenhagen, Denmark [24] Department of Viruses, Hormones and Cancer, Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark [25] Department of Statistical Science, Duke University [26] International Hereditary Cancer Center, Department Of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland [27] Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Cancer Centre and Institute of Oncology, Warsaw, Poland [28] Department of Health Research and Policy, Stanford University School of Medicine, Stanford, USA [29] Samuel Lunenfeld Research Institute, 60 Murray Street, Toronto, ON, Canada [30] Roswell Park Cancer Centre, Buffalo, NY, USA [31] Center for Research in Women's Health, 790 Bay St., Toronto, ON, Canada [32] Yale University, 60 College Street, New Haven, CT, USA [33] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, VIC 3010, Australia [34] University of Southern California, Keck School of Medicine, Los Angeles, California, USA [35] Department of Obstetrics and Gynaecology, University of Ulm, Ulm, Germany [36] Department of Obstetrics and Gynecology, University North Carolina, Chapel Hill, North Carolina, USA [37] Queensland Institute of Medical Research and Peter MacCallum Cancer Center [38] Section of Cancer Genetics, Institute of Cancer Research, Sutton, UK [39] Population and Functional Genetics Lab, Wellcome Trust Centre for Human Genetics, Oxford, UK [40] The University of Texas, School of Public Health, Houston, TX, USA [41] University of California at Los Angeles, David Geffen School of Medicine, Division of Hematology and Oncology [42] CR-UK Genetic Epidemiology Unit, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK

Abstract

Epithelial ovarian cancer has a major heritable component, but the known susceptibility genes explain less than half the excess familial risk[1]. We performed a genome wide association study (GWAS) to identify common ovarian cancer susceptibility alleles. We evaluated 507,094 SNPs genotyped in 1,817 cases and 2,353 controls from the UK and ~2 million imputed SNPs. We genotyped the 22,790 top ranked SNPs in 4,274 cases and 4,809 controls of European ancestry from Europe, USA and Australia. We identified 12 SNPs at 9p22 associated with disease risk ($P<10^{-8}$). The most significant SNP (rs3814113; $P = 2.5 \times 10^{-17}$) was genotyped in a further 2,670 ovarian cancer cases and 4,668 controls confirming its association (combined data odds ratio = 0.82 95% CI $0.79 - 0.86$, $P_{-trend} = 5.1 \times 10^{-19}$). The association differs by histological subtype, being strongest for serous ovarian cancers (OR 0.77 95% CI $0.73 - 0.81$, $P_{trend} = 4.1 \times 10^{-21}$).

Women with a first-degree relative diagnosed with epithelial ovarian cancer have a three-fold increased risk of developing the disease[2]. Environmental and genetic factors contribute to this increased risk, but studies of twins suggest that genetic factors are more important. *BRCA1* and *BRCA2* mutations confer high risk of ovarian cancer and are responsible for most families with three or more ovarian cancer cases. They account for less than half the excess familial risk[3, 4] and it is likely that the residual risk is due to a combination of common and/or rare alleles that confer moderate to low penetrance susceptibility[5].

Many recent studies have reported the identification of common alleles that confer low-penetrance susceptibility to common cancers including breast, prostate and colorectal cancers and melanoma6[−12]. These studies all used a genome wide association study (GWAS) design in which the genotype frequencies of hundreds of thousands of single nucleotide polymorphisms (SNPs) distributed throughout the genome are compared between large numbers of cases and unaffected controls. In the current study, we conducted a three-stage GWAS to identify alleles associated with variation in the risks of invasive epithelial ovarian cancer (Table 1). In the first stage, we used the Illumina Infinium 610K array to genotype 620,901 SNPs that report on genetic variation across the genome in 1,890 ovarian cancer cases recruited throughout the UK (see Supplementary Table 1 online for details). We enriched for likely genetic heritability by including 47 cases from families with two or more ovarian cancer cases that had been screened negative for *BRCA1* or *BRCA2* mutations. After excluding 73 cases that failed to meet genotyping quality control criteria, the genotype frequencies for 1,817 cases were compared with the genotypes of 2,353 UK controls that had been analysed using the similar 550k array on the same genotyping platform as part of a GWAS for other phenotypes (Supplementary Table 1 online). All the subjects analyzed in stage 1 were of European ancestry. After excluding 80,327 SNPs that were not genotyped on controls or failed to genotype on cases, a total of 507,094 SNPs with a minor allele frequency (MAF) of at least 1% in controls passed genotype quality control criteria. We also evaluated an additional ~2million SNPs with genotypes imputed using the phase2 Hapmap data (CEU).

Supplementary figure 1a shows the quantile-quantile (Q-Q) plot of the distribution of test statistics for comparison of genotype frequencies in cases versus controls (1 degree freedom (d.f) Cochran-Armitage trend test) for SNPs genotyped in stage 1. There was little evidence of any general inflation of the test statistics (estimated inflation factor $\lambda_{1000} = 1.026$ based on the bottom 90% of the distribution13). This is consistent with the population structure in the UK and for other GWAS that have used UK populations[9, 14].

In a second stage, 23,590 SNPs were genotyped in 10 studies comprising 4,964 cases of invasive epithelial ovarian cancer and 5,379 controls (Table 1) using the Illumina iSelect platform. We selected 22,790 SNPs based on the lowest P-trend in tests for association with ovarian cancer risk from the ~2.5 million genotyped and imputed SNPs, and 800 SNPs that reported on ancestry. Data for 273 subjects were excluded because they did not meet quality control criteria; thus, genotyping data were available on 4,833 cases and 5,237 controls. Five hundred and fifty nine cases and 428 controls in stage 2 were of non-European ancestry and excluded from the main analyses. Supplementary figure 1B shows the Q-Q plot for SNPs genotyped in stage 2. Again, there was little evidence of any general inflation of the test statistics for stage 2 (estimated inflation factor $\lambda_{1000} = 1.005$). In the combined analysis of stage 1 and stage 2 data, we identified 12 SNPs associated with ovarian cancer risk with $P < 10^{-8}$ (Supplementary Table 2 online). All 12 SNPs were located in the same region on chromosome 9 (9p22.2). The strongest association was for rs3814113, which was the only SNP retained in a multivariate logistic regression model. We genotyped rs3814113 in an additional 3,089 cases of invasive epithelial ovarian cancer and 5,340 controls from ten studies that are part of the Ovarian Cancer Association Consortium (OCAC) (Supplementary Table 1 online). Four hundred and nineteen cases and 672 controls were of self-reported non-European ancestry and excluded from the analysis. These additional data reinforced the evidence of association for rs3814113; P-trend = 5.1×10^{-19} based on data from all three stages (Table 2).

rs3814113 was associated with a decrease in the risk of ovarian cancer in carriers of the minor allele (per minor allele odds ratio (OR) = 0.82 95% confidence interval (CI) 0.79- 0.86). The effect size was similar in stages 1 and 2, but slightly smaller in stage 3. There was no heterogeneity in the OR estimates amongst studies for any stages (Table 2 and in Figure 1A). Based on an odds ratio of 0.82 and minor allele frequency of 0.32, this locus explains

approximately 0.7% of the polygenic component of ovarian cancer risk. This estimate is also based on the assumption that the known high penetrance genes explain 40% of the excess familial risk and the unexplained component is polygenic. There was no significant difference between the risk of ovarian cancer in subjects of European and non-European ancestry (P = 0.83). However, the per-minor allele risk was slightly attenuated and not significant in the subjects of non-European ancestry (OR = 0.89, 95% CI 0.78–1.01, P = 0.077) (Supplementary Table 3 online).

We also evaluated the association for rs3814113 with ovarian cancer risk after stratifying cases by histological subtype. The strength of association increased when serous cases (n = 4,847), the most common histological subtype, were considered alone (OR = 0.77 95% CI 0.73–0.81, P-trend = 4.1×10^{-21}) (Supplementary Table 4 online and Figure 1B). When the analysis was restricted to serous cases the effects were similar between European and non-European subject (non-European OR = 0.79, 95% CI 0.66–0.94, P = 0.007). We only detected marginal evidence of association for rs3814113 in 1,320 cases diagnosed with endometrioid ovarian cancer (OR = 0.86 95% CI 0.79–0.94, P-trend = 0.001), and no association for patients with mucinous (n = 626) or clear cell (n = 628) ovarian cancer (Supplementary Table 4 online). However, the small numbers of mucinous or clear cell cases limited the power to detect modest effects. Compared to the non-serous subtypes, the ovarian cancer risk in serous cases was significantly lower (P= 7.8×10^{-5}). There was also some suggestion of a bigger effect in older women (P-trend = 0.006 for ovarian cancer overall and P-trend = 0.044 for serous type ovarian cancer) (Supplementary Table 5 online). There was no significant difference in genotype frequency for cases reporting a family history of ovarian cancer compared with cases with negative family history of ovarian cancer (P = 0.59)

All 12 SNPs in the region that were identified after stage 2 were in the same LD block and were associated with a decreased risk of ovarian cancer, perhaps suggesting that susceptibility is driven by a single correlated variant within the region (Figure 2), although it is possible that there are multiple independent SNPs all correlated with the best markers. rs3814113 may be the causal SNP on 9p22.2 or it may be a marker in linkage disequilibrium (LD) with the true functional variant or haplotype. Neither rs3814113 nor the highly correlated SNP, rs4445329, ($r^2 > 0.99$) are located within an open reading frame or an intronic region of any gene. The nearest genes are *BNC2* (basonuclin 2)*, CNTLN* (centlein, a centrosomal protein) and hypothetical gene *LOC648570*. rs3814113 is ~44kb upstream of *BNC2*, ~128kb upstream of *LOC648570* and ~220kb downstream of *CNTLN*. Eight of the associated SNPs in the region are located within intron 2 of *BNC2* (Figure 2). *BNC2* encodes DNA-binding zinc-finger protein that is highly conserved across vertebrates suggesting it is an important regulatory protein for DNA transcription[15]. The gene exhibits extensive transcriptional variability; it has six promoters and has the potential to generate up to 90,000 mRNA isoforms encoding more than 2,000 different proteins[16]. The Genevar project provides data on gene expression of *BNC2* in lymphocyte derived cell lines from the CEU population based on nine probes in the region[17]. There was no association for any of these probes and genotype at the top 12 SNPs. Also none of the top 12 SNPs appear to be near predictable or known enhancer binding sites or splice sites using PupaSNP (http://www.pupasnp.org/). *BNC2* is highly expressed in reproductive tissues (ovary and testis) and may play a role in the differentiation of spermatoza and oocytes[18]. There is little evidence of a role for *BNC2* in cancer development, although there is a report of 7~9 fold up-regulated expression in basal cell carcinoma compared to normal basal cells[19]. Resequencing of the 9p22.2 region and further genotyping in ovarian cancer cases and controls will be needed to clarify the likely causal variant(s).

We found no additional susceptibility loci reaching genome wide significance ($P<10^{-8}$) in this study. Positive associations have previously been reported in candidate gene studies, but none of these reached genome-wide significance. Supplementary Table 6 shows some of the more

notable published associations[20–29] together with the results for the same SNPs from this study. In general the reported associations and our results are consistent, but none of the associations in this study ranked highly enough for the SNP to be included in the Stage 2 genotyping. The power to have identified rs3814113 depends on the true relative risk. Assuming a relative risk per-minor allele of 0.82 (combined data estimate), power was > 90% at genome-wide significance, suggesting that common alleles of larger effect are unlikely to have been missed. However, if the true per-allele relative risk is 0.88 (stage 3 estimate) then the power was only 20% suggesting that susceptibility alleles with more modest effects remain to be identified. Furthermore, power to detect less common alleles will be limited unless the effect on risk is greater. For example, we have 90 percent power at genome-wide significance to detect a risk allele with 5 percent frequency that confers a relative risk of 1.44. This would be consistent with the findings for other common cancer types (e.g. breast cancer and prostate cancers)[29]. In addition, disease heterogeneity may have limited our power to identify additional susceptibility alleles. In the primary analysis, we considered ovarian cancer as a single disease phenotype, but the effects of association for rs3814113 varied when cases were stratified by histological subtype. Different subtypes of ovarian cancer have different biological properties and this finding supports previous studies that suggest susceptibility due to germline genetic variation may be subtype specific. For example, serous and endometrioid cancer are relatively more common in *BRCA1* and *BRCA2* mutation carriers whereas mutations in the DNA mismatch repair genes are more frequently associated with mucinous ovarian cancers[30–32]. Additional ovarian cancer susceptibility loci may exist that are associated with specific histological or molecular subtypes[33]. However, the power to identify alleles for the rarer subtypes (endometrioid, mucinous and clear cell ovarian cancers) was limited by the numbers of cases in the study. It is likely that pooling of data from multiple ovarian cancer GWAS will enable additional susceptibility alleles for invasive ovarian cancer in general and for specific sub-types to be identified.

The identification of common ovarian cancer susceptibility variants may have clinical implications in the future for identifying patients at greatest risk of the disease. Survival rates in patients diagnosed with ovarian cancer are poor - approximately 70% of patients are diagnosed with late stage disease and less than 40% of these cases survive more than 5 years after their diagnosis. The efficacy of using multi-modal approaches to early detection of the disease are limited, but may be improved by using genetic risk profiling to identify a subset of the population that would benefit most from earlier disease detection. The benefits of a similar approach have recently been modelled in breast cancer[5]. Identifying genetic variants that cause ovarian cancer may also improve our understanding of the underlying biology of ovarian cancer, potentially leading to the development of more effective, individualised therapies. For example, the identification of the highly penetrant susceptibility genes *BRCA1* and *BRCA2*, and their subsequent functional characterisation has since led the to development of a potential novel therapy for patients deficient in BRCA1/2 function based on inhibition of the poly (ADP-ribose) polymerase PARP DNA repair pathway[34,35]. The 9p22.2 region is the first common susceptibility locus for ovarian cancer to be established. Understanding the mechanisms by which this susceptibility is mediated should improve our understanding of the biology of ovarian cancer, and may lead to new approaches to treat or prevent the disease.

METHODS

Study design

The ovarian cancer case-control studies that participated in stages 1, 2 and 3 are summarized in Supplementary Table 1 online. Stage 1 comprised invasive epithelial ovarian cancer cases from UK and genotype data of UK controls from GWAS of other phenotypes. Stage 2 comprised 10 studies from the OCAC. Stage 3 comprised 10 additional studies from the OCAC.

For all studies we have data on disease status, age at diagnosis and date of blood draw, self-reported ethnic group and histological subtype. All but 5 studies provided information of reported first-degree family history of ovarian cancer.

Genotyping

Genotyping for stage 1 cases was conducted using the Illumina Infinium 610K array at Illumina Corporation. Existing data from two sets of controls, genotyped on the Infinium 550k array, were used in stage 1 analyses: the Welcome Trust Case-Control Consortium 1958 birth cohort 14, and a national colorectal control study[36] using Illumina platform Hap550 array. All cases were from the UK and confirmed as invasive epithelial ovarian cancer. Quality control criteria were applied separately to the cases and each control sets because they were genotyped separately. SNPs were excluded if (1) they deviated from hardy-Weinberg equilibiurm (HWE) at $P < 10^{-4}$ or (2) had a MAF < 1%, or (3) MAF was between 1% and 5% and call rate < 99% or (4) MAF > 5% and call rate < 95%. We also rejected SNPs if a test for trend by genotype between the two control sets was significant at $P < 10^{-4}$. This led to 33,479 SNPs being excluded and 507,094 SNPs passing QC. Genotyping the 10 studies in stage 2 was conducted using an Illumina iSelect array at Illumina Corportation. We excluded SNPs (n=1,635) for the stage 2 data if the sum of the test statistics for deviation from HWE for the 10 studies was significant at $p < 10^{-5}$ or if the SNPs had a call rate of <95% or if the MAF < 0.5%. A total of 21,955 SNPs were available for data analysis in stage 2.

We utilized only samples with called genotypes on at least 80% of SNPs. Seventy-three samples were excluded from Stage 1 and 273 from Stage 2. Nineteen samples were included as duplicates in stage 1 and genotype concordance rate for these pairs was 99.99%. One hundred and twenty-two samples were included as duplicates in stage 2 and duplicate concordance rate was 99.99%. Six studies in stage 3 were genotyped for rs3814113 by Taqman using ABI Prism 7900HT sequence detection system at each laboratory. For three studies (TOR, NCO, MAY) genotype data were available from an independent, ongoing GWAS study that is also used the Illumina Infinium 610K platform. Genotyping and QC was performed at Mayo Clinic genotyping shared resources. For Taqman genotyping quality control, we compared genotype call rates and concordance by study and overall. We used the following criteria as a measure of acceptable genotyping: (1) > 3% sample duplicates included; (2) concordance rate for the duplicates ≥ 98%; (3) overall call rate (by study) > 95%; (4) call rates > 90% for each individual 384-well plate and (5) no deviation from HWE in controls (P>=0.05). Genotyping consistency across laboratories using Taqman was also evaluated by genotyping a common panel of CEPH-Utah trios including 90 individual DNA samples, 5 duplicate samples and 1 negative control (http://ccr.coriell.org/). The concordance of genotyping results between the centres was required to be greater than 98% in order for the genotype data to be included. The genotyping results from all studies in the stage 3 met the above criteria and were included in the final analysis.

Statistics

Population stratification—For the stage 1 samples, we used the program LAMP[37] to assign intercontinental ancestry based on the Hapmap (release #22) genotype frequency data for European, African and Asian populations. Samples with less than 90% European ancestry were excluded from the analysis (n=73). For the stage 2 data, 800 SNPs that are known to be predictive of ancestry ("Ancestry Informative Markers") (AIMs) were genotyped. We again used LAMP and the Hapmap data (release #23) on European (CEU), African American (ASW), East Asian (JPT-CHB-CHD), Mexican (MEX) and Indian (GIH) populations to estimate ancestry. Subjects with less than 90 percent European ancestry were excluded from the main analyses (n= 987). We then used the AIMs to calculate principal components for the subjects of European ancestry. The first principal component explained 0.42 percent of the variability

and was included as a covariate in subsequent association analyses. Subsequent principal components were not included as they and explained less variability and there was little difference in their Eigenvalues.

Imputation—We imputed missing genotype data for all the common variants in the Hapmap data for two reasons. Firstly the stage 1 cases and controls were genotyped using slightly different SNP sets and different SNPs may have failed QA in different sample sets. Secondly, imputing SNPs that have not been genotyped increases genome coverage and may improve power. We used an in-house method that combines the features of fastPHASE [38] and IMPUTE [39] to impute the ungenotyped or missing SNPs, utilising the phase2 Hapmap data (CEU) which contains phased haplotypes for 60 individuals on 2.5 million SNPs. For each imputed genotype the expected number of minor alleles carried was estimated (weights). Genotyped SNPs were assigned weights of 0, 1 or 2 (actual number of minor alleles carried). We estimated the accuracy of imputation by calculating the estimated r^2 between the imputed and actual SNP [40]. SNPs with $r^2 < 0.64$ were excluded (n = 152,401) leaving a total of 2,563,972 SNPs for stage 1 analysis.

Statistical Tests of Association—We used logistic regression to test for association between genotype and case -control status using the imputed weights and the ethnicity estimates as covariates. SNPs were selected for replication in Stage 2 based on the weighted ranked test statistics from the Stage 1 analysis. The weights were based on whether the SNP was genotyped or imputed ($I_{genotype} = 1$ if genotyped,, 0 otherwise), the accuracy of imputation (r^2), the association statistic (T) and the design score (s) using somewhat arbitrary formula (($1 + 0.1 I_{genotype}$)$*T*r^2 + 2s$). SNPs that were correlated with a higher ranked SNP with $r^2 > 0.8$ were excluded except for those correlated with the 1000 highest ranked SNPs. Using this method we took forward 22,790 SNPs as candidates for association of which 5,380 were purely imputed (not-genotyped). For the stage 2 analysis, we tested for association by performing logistic regression using the imputed values from stage 1, combined with the genotyping results for stage 2. We corrected for the first principal component and the ethnicity estimates in the second stage analysis (all subjects in the first stage were selected to be European) and stratified by study using a Wald test. We corrected for ethnicity in stage 3 using self reported ethnicity and stratified by study. A subgroup analysis was used to compare genotype-specific risks by disease subgroup with controls. The effect of age group, family history and population of origin (European and non-European) was assessed similarly. Modification by these sub-groups was tested by fitting a SNP by subgroup interaction term in a logistic regression model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Pharoah PD, Ponder BA. The genetics of ovarian cancer. Best Pract Res Clin Obstet Gynaecol 2002;16:449–468. [PubMed: 12413928]

2. Stratton JF, Pharoah P, Smith SK, Easton D, Ponder BA. A systematic review and meta-analysis of family history and risk of ovarian cancer. Br J Obstet Gynaecol 1998;105:493–499. [PubMed: 9637117]

3. Antoniou AC, Easton DF. Risk prediction models for familial breast cancer. Future Oncol 2006;2:257–274. [PubMed: 16563094]

4. Gayther SA, et al. The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer-susceptibility genes. Am J Hum Genet 1999;65:1021–1029. [PubMed: 10486320]

5. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358:2796–2803. [PubMed: 18579814]

6. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–1093. [PubMed: 17529967]

7. Thomas G, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310–315. [PubMed: 18264096]

8. Brown KM, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 2008;40:838–840. [PubMed: 18488026]

9. Eeles RA, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008;40:316–321. [PubMed: 18264097]

10. Gudbjartsson DF, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. Nat Genet 2008;40:886–891. [PubMed: 18488027]

11. Broderick P, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1315–1317. [PubMed: 17934461]

12. Yeager M, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007;39:645–649. [PubMed: 17401363]

13. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008;17:R122–R128. [PubMed: 18852200]

14. Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678. [PubMed: 17554300]

15. Vanhoutteghem A, Djian P. Basonuclin 2: an extremely conserved homolog of the zinc finger protein basonuclin. Proc Natl Acad Sci U S A 2004;101:3468–3473. [PubMed: 14988505]

16. Vanhoutteghem A, Djian P. The human basonuclin 2 gene has the potential to generate nearly 90, 000 mRNA isoforms encoding over 2000 different proteins. Genomics 2007;89:44–58. [PubMed: 16942855]

17. Stranger BE, et al. Population genomics of human gene expression. Nat Genet 2007;39:1217–1224. [PubMed: 17873874]

18. Romano RA, Li H, Tummala R, Maul R, Sinha S. Identification of Basonuclin2, a DNA-binding zinc-finger protein expressed in germ tissues and skin keratinocytes. Genomics 2004;83:821–833. [PubMed: 15081112]

19. O'Driscoll L, et al. Investigation of the molecular profile of basal cell carcinoma using whole genome microarrays. Mol Cancer 2006;5:74. [PubMed: 17173689]

20. Song H, et al. Association between invasive ovarian cancer susceptibility and 11 best candidate SNPs from breast cancer genome-wide association study. Hum Mol Genet 2009;18:2297–2304. [PubMed: 19304784]

21. Quaye L, et al. The effects of common genetic variants in oncogenes on ovarian cancer survival. Clin Cancer Res 2008;14:5833–5839. [PubMed: 18794094]

22. Pearce CL, et al. Validating genetic risk associations for ovarian cancer through the international Ovarian Cancer Association Consortium. Br J Cancer. 2009

23. Schildkraut JM, et al. Single nucleotide polymorphisms in the TP53 region and susceptibility to invasive epithelial ovarian cancer. Cancer Res 2009;69:2349–2357. [PubMed: 19276375]

24. Ramus SJ, et al. Consortium analysis of 7 candidate SNPs for ovarian cancer. Int J Cancer 2008;123:380–388. [PubMed: 18431743]

25. Ghoussaini M, et al. Multiple Loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst 2008;100:962–966. [PubMed: 18577746]

26. Song H, et al. Association Study of Prostate Cancer Susceptibility Variants with Risks of Invasive Ovarian, Breast, and Colorectal Cancer. Cancer Res 2008;68:8837–8842. [PubMed: 18974127]

27. Gayther SA, et al. Tagging single nucleotide polymorphisms in cell cycle control genes and susceptibility to invasive epithelial ovarian cancer. Cancer Res 2007;67:3027–3035. [PubMed: 17409409]

28. Song H, et al. Tagging Single Nucleotide Polymorphisms in the BRIP1 Gene and Susceptibility to Breast and Ovarian Cancer. PLoS ONE 2007;2:e268. [PubMed: 17342202]

29. Easton DF, Eeles RA. Genome-wide association studies in cancer. Hum Mol Genet 2008;17:R109–R115. [PubMed: 18852198]

30. Rubin SC, et al. Clinical and pathological features of ovarian cancer in women with germ-line mutations of BRCA1. N Engl J Med 1996;335:1413–1416. [PubMed: 8875917]

31. Lakhani SR, et al. Pathology of ovarian cancers in BRCA1 and BRCA2 carriers. Clin Cancer Res 2004;10:2473–2481. [PubMed: 15073127]

32. Domanska K, Malander S, Masback A, Nilbert M. Ovarian cancer at young age: the contribution of mismatch-repair defects in a population-based series of epithelial ovarian cancer before age 40. Int J Gynecol Cancer 2007;17:789–793. [PubMed: 17343610]

33. Tothill RW, et al. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 2008;14:5198–5208. [PubMed: 18698038]

34. Lord CJ, Ashworth A. Targeted therapy for cancer using PARP inhibitors. Curr Opin Pharmacol 2008;8:363–369. [PubMed: 18644251]

35. Drew Y, Calvert H. The potential of PARP inhibitors in genetic breast and ovarian cancers. Ann N Y Acad Sci 2008;1138:136–145. [PubMed: 18837894]

36. Houlston RS, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426–1435. [PubMed: 19011631]

37. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. Am J Hum Genet 2008;82:290–303. [PubMed: 18252211]

38. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629–644. [PubMed: 16532393]

39. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 2007;39:906–913. [PubMed: 17572673]

40. Udler MS, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. Hum Mol Genet 2009;18:1692–1703. [PubMed: 19223389]
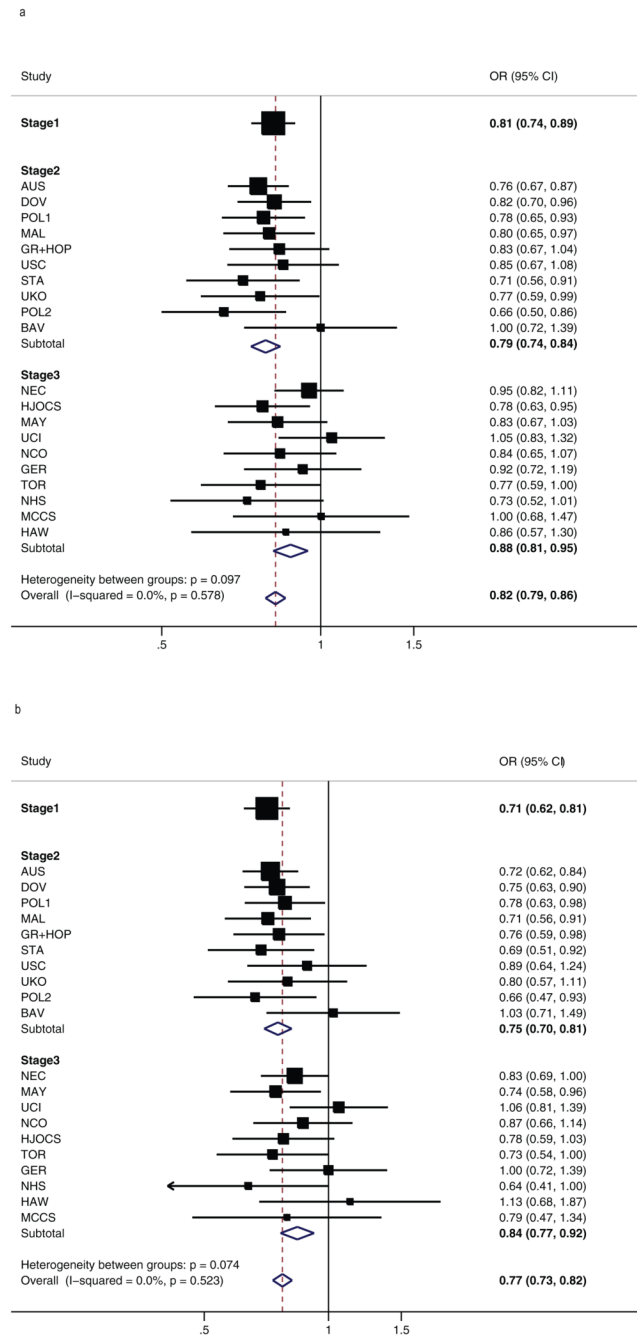
a

Study		OR (95% CI)
Stage1		**0.81 (0.74, 0.89)**
Stage2		
AUS		0.76 (0.67, 0.87)
DOV		0.82 (0.70, 0.96)
POL1		0.78 (0.65, 0.93)
MAL		0.80 (0.65, 0.97)
GR+HOP		0.83 (0.67, 1.04)
USC		0.85 (0.67, 1.08)
STA		0.71 (0.56, 0.91)
UKO		0.77 (0.59, 0.99)
POL2		0.66 (0.50, 0.86)
BAV		1.00 (0.72, 1.39)
Subtotal		**0.79 (0.74, 0.84)**
Stage3		
NEC		0.95 (0.82, 1.11)
HJOCS		0.78 (0.63, 0.95)
MAY		0.83 (0.67, 1.03)
UCI		1.05 (0.83, 1.32)
NCO		0.84 (0.65, 1.07)
GER		0.92 (0.72, 1.19)
TOR		0.77 (0.59, 1.00)
NHS		0.73 (0.52, 1.01)
MCCS		1.00 (0.68, 1.47)
HAW		0.86 (0.57, 1.30)
Subtotal		**0.88 (0.81, 0.95)**
Heterogeneity between groups: p = 0.097		
Overall (I-squared = 0.0%, p = 0.578)		**0.82 (0.79, 0.86)**

.5 1 1.5

b

Study		OR (95% CI)
Stage1		**0.71 (0.62, 0.81)**
Stage2		
AUS		0.72 (0.62, 0.84)
DOV		0.75 (0.63, 0.90)
POL1		0.78 (0.63, 0.98)
MAL		0.71 (0.56, 0.91)
GR+HOP		0.76 (0.59, 0.98)
STA		0.69 (0.51, 0.92)
USC		0.89 (0.64, 1.24)
UKO		0.80 (0.57, 1.11)
POL2		0.66 (0.47, 0.93)
BAV		1.03 (0.71, 1.49)
Subtotal		**0.75 (0.70, 0.81)**
Stage3		
NEC		0.83 (0.69, 1.00)
MAY		0.74 (0.58, 0.96)
UCI		1.06 (0.81, 1.39)
NCO		0.87 (0.66, 1.14)
HJOCS		0.78 (0.59, 1.03)
TOR		0.73 (0.54, 1.00)
GER		1.00 (0.72, 1.39)
NHS		0.64 (0.41, 1.00)
HAW		1.13 (0.68, 1.87)
MCCS		0.79 (0.47, 1.34)
Subtotal		**0.84 (0.77, 0.92)**
Heterogeneity between groups: p = 0.074		
Overall (I-squared = 0.0%, p = 0.523)		**0.77 (0.73, 0.82)**

.5 1 1.5

Figure 1.
Genotype-specific risks of SNP rs3814113 for ovarian cancer by stage and by study - (a) all epithelial ovarian cancer cases, (b) serous ovarian cancer cases only. Results are based on analyses restricted to subjects of European ancestry.
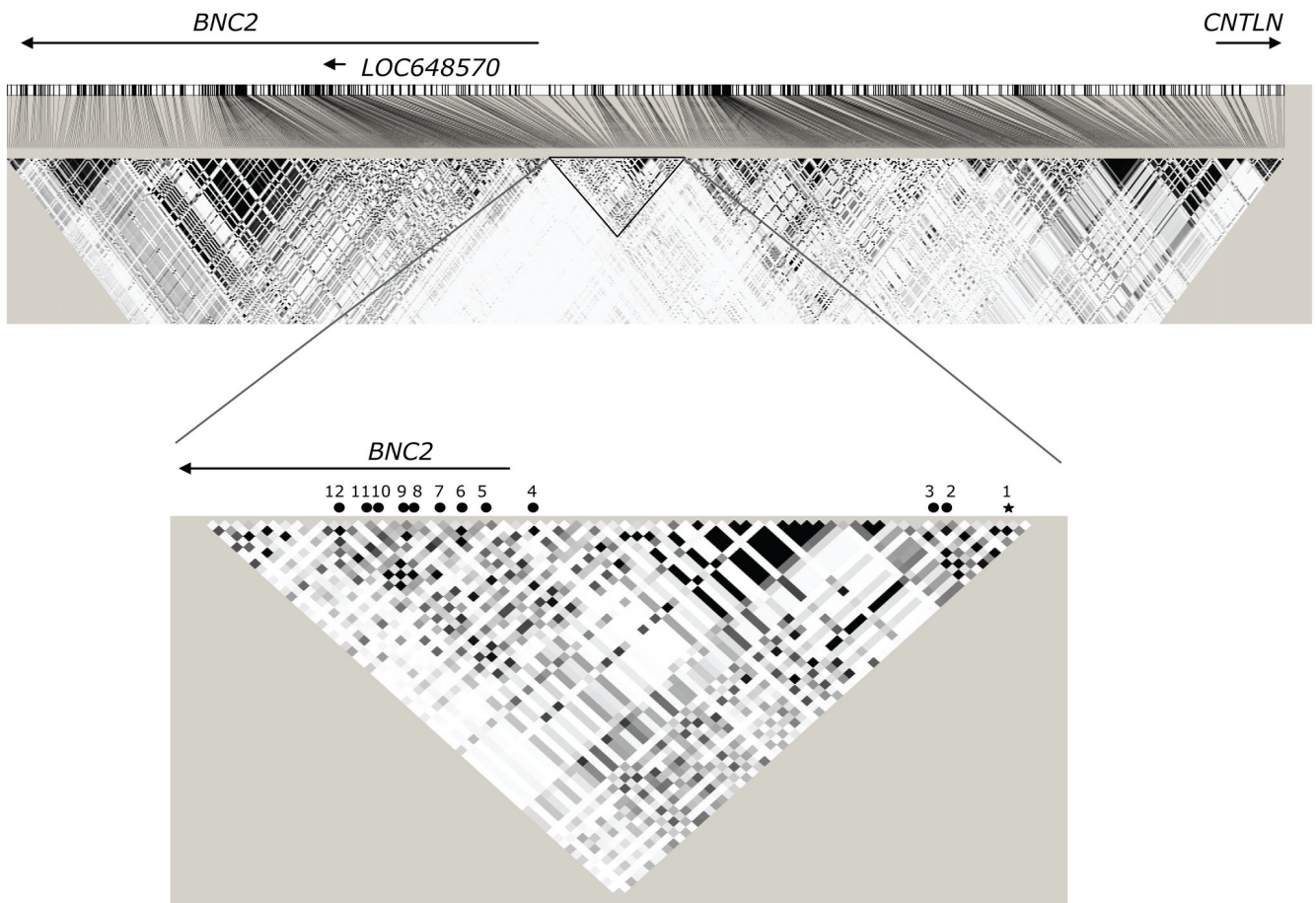
Figure 2.
Map of associated regions on 9p22.2 (nucl. 16655021-17155021). The top panel shows the LD blocks around rs3814113 (±250 kb) for SNPs with MAF≥ 0.05 based on Hapmap CEU data. Squares on the LD block indicate the correlation (r^2) between SNPs on a greyscale (darker squares = higher correlations). Approximate location of *BNC2* (nucl. 16860786-16399502), *CNTLN* (nucl. 17125038-17493915) and *LOC648570* (nucl. 16777433-16775576) were inferred from the NCBI reference assembly (solid line). The lower panel shows the highlighted LD block in greater detail. The star indicates the location of the most strongly associated SNP rs3814113. The black circles show the position of the other 11 SNPs that reached genome wide significance. The numbers refer to the SNP identifiers in Supplementary Table 2.

Table 1

Summary of study design.

Stage	Study population	Genotyping method	No. SNPs genotyped	No. SNPs passed QC	No. SNPs imputed	Number of Subjects [1]	
						Cases	Controls
Stage 1	UK	Illumina 610K	620,901	507,094	2,056,878	1,890 (73)	2,353
Stage 2	Europe, USA, Australia	Illumina iSelect	23,590	21,955	0	4,964 (131)	5,379 (142)
Stage 3	Europe, USA, Canada	Taqman or Illumina	1	1	0	**3,089**	**5,340**
	Total					**9,943 (204)**	**13,072 (142)**

[1] The numbers of subjects that were excluded from analysis due to failing genotyping quality control criteria are presented in brackets.

Table 2

Estimated odds ratio and tests of association for rs3814113 at chromosome 9 (9p22.2) based on combined data for subjects of European ancestry.

Tumor subtype	Stage	No. of Cases/Controls	MAF[a] Case/Control	Per-allele OR[b] [95% CI]	Heterozygote[c] OR [95% CI]	Homozygote[d] OR [95% CI]	P-trend
All Cases	Stage 1	1817/2354	0.27/0.32	0.81 (0.74–0.89)	0.76 (0.67–0.86)	0.73 (0.58–0.91)	2.04×10^{-5}
	Stage 2	4274/4809	0.28/0.33	0.79 (0.74–0.84)	0.80 (0.76–0.93)	0.70 (0.63–0.77)	1.98×10^{-13}
	Stage 3	2670/4668	0.29/0.32	0.88 (0.81–0.94)	0.84 (0.77–0.93)	0.80 (0.68–0.95)	4.00×10^{-4}
	Combined	**8761/11831**	**0.28/0.32**	**0.82 (0.79–0.86)**	**0.80 (0.75–0.85)**	**0.70 (0.63–0.77)**	**5.10×10^{-19}**
Serous Cases	Stage 1	837/2354	0.25/0.32	0.71 (0.62–0.81)	0.70 (0.59–0.82)	0.52 (0.38–0.72)	1.44×10^{-7}
	Stage 2	2483/4809	0.27/0.33	0.75 (0.70–0.81)	0.76 (0.68–0.84)	0.57 (0.47–0.68)	5.60×10^{-13}
	Stage 3	1527/4668	0.28/0.32	0.83 (0.76–0.92)	0.81(0.71–0.92)	0.74 (0.60–0.92)	2.07×10^{-4}
	Combined	**4847/11831**	**0.27/0.32**	**0.77 (0.73–0.81)**	**0.76 (0.71–0.82)**	**0.61 (0.54–0.69)**	**4.10×10^{-21}**

[a]Minor allele frequency in populations with European ancestry in the studies

[b]Odds ratios (OR) and 95% confidence interval per copy of minor allele

[c]OR for heterozygotes compared with common homozygotes

[d]OR for rare homozygotes compared with common homozygote. Data highlighted with bold text are GWAS significant (P<5.0×10^{-8})