Genome-Wide Association Study for Ovarian Cancer Susceptibility Using Pooled DNA

Yi Lu,[1] Xiaoqing Chen,[1] Jonathan Beesley,[1] Sharon E. Johnatty,[1] Anna deFazio,[2,3] Australian Ovarian Cancer Study (AOCS) Study Group, Sandrina Lambrechts,[4] Diether Lambrechts,[5,6] Evelyn Despierre,[4] Ignace Vergotes,[4] Jenny Chang-Claude,[7] Rebecca Hein,[7] Stefan Nickels,[7] Shan Wang-Gohrke,[8] Thilo Dörk,[9] Matthias Dürst,[10] Natalia Antonenkova,[11] Natalia Bogdanova,[11,12] Marc T. Goodman,[13] Galina Lurie,[13] Lynne R. Wilkens,[13] Michael E. Carney,[14] Ralf Butzow,[15] Heli Nevanlinna,[15] Tuomas Heikkinen,[15] Arto Leminen,[15] Lambertus A. Kiemeney,[16,17,18] Leon F.A.G. Massuger,[19] Anne M. van Altena,[19] Katja K. Aben,[17,18] Susanne Krüger Kjaer,[20] Estrid Høgdall,[20] Allan Jensen,[21] Angela Brooks-Wilson,[22,23] Nhu Le,[24] Linda Cook,[25] Madalene Earp,[23] Linda Kelemen,[26] Douglas Easton,[27] Paul Pharoah,[27] Honglin Song,[27] Jonathan Tyrer,[27] Susan Ramus,[28] Usha Menon,[29] Alexandra Gentry-Maharaj,[29] Simon A. Gayther,[28] Elisa V. Bandera,[30,31] Sara H. Olson,[31] Irene Orlow,[31] Lorna Rodriguez-Rodriguez,[30] Stuart Macgregor,[1]* and Georgia Chenevix-Trench[1]*

[1] Queensland Institute of Medical Research, Brisbane, Australia
[2] Department of Obstetrics and Gynaecology, University of Sydney, Sydney, Australia
[3] Westmead Institute for Cancer Research, Westmead Hospital, Sydney, Australia
[4] Division of Gynaecological Oncology, Department of Obstetrics and Gynaecology, University Hospital Leuven, University of Leuven, Leuven, Belgium
[5] Vesalius Research Center, VIB, Leuven, Belgium
[6] Vesalius Research Center, University of Leuven, Leuven, Belgium
[7] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[8] Department of Obstetrics and Gynecology, University of Ulm, Ulm, Germany
[9] Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
[10] Department of Gynaecology, Friedrich Schiller University, Jena, Germany
[11] Byelorussian Institute for Oncology and Medical Radiology, Aleksandrov NN, Minsk, Belarus
[12] Clinics of Radiation Oncology, Hannover Medical School, Hannover, Germany
[13] Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
[14] Department of Obstetrics and Gynecology, John A Burns School of Medicine, University of Hawaii, Honolulu, HI, USA
[15] Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland
[16] Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[17] Department of Epidemiology, Biostatistics and HTA, Nijmegen Medical Centre, Radboud University, Nijmegen, The Netherlands
[18] Comprehensive Cancer Center, Nijmegen, The Netherlands
[19] Department of Gynaecology, Nijmegen Medical Centre, Radboud University, Nijmegen, The Netherlands
[20] Department of Viruses, Hormones and Cancer, Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark
[21] Department of Gynecology, Juliane Marie Centre, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
[22] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada
[23] Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada
[24] Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, Canada
[25] Division of Epidemiology and Biostatistics, University of New Mexico, Albuquerque, NM, USA
[26] Alberta Health Services — Cancer Care, Calgary, Alberta, Canada
[27] Departments of Oncology and Public Health and Primary Care, University of Cambridge, Cambridge, UK
[28] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[29] Gynaecological Oncology Unit, UCL EGA Institute for Women's Health, University College London, London, UK
[30] The Cancer Institute of New Jersey, New Brunswick, NJ, USA
[31] Memorial Sloan-Kettering Cancer Center, New York, NY, USA
* These two authors contributed equally.

Recent Genome-Wide Association Studies (GWAS) have identified four low-penetrance ovarian cancer susceptibility loci. We hypothesized that further moderate- or low-penetrance variants exist among the subset of single-nucleotide polymorphisms (SNPs) not well tagged by the genotyping arrays used in the previous studies, which would account for some of the remaining risk. We therefore conducted a time- and cost-effective stage 1 GWAS on 342 invasive serous cases and 643 controls genotyped on pooled DNA using the high-density Illumina 1M-Duo array. We followed up 20 of the most significantly associated SNPs, which are not well tagged by the lower density arrays used by the published GWAS, and genotyping them on individual DNA. Most of the top 20 SNPs were clearly validated by individually genotyping the samples used in the pools. However, none of the 20 SNPs replicated when tested for association in a much larger stage 2 set of 4,651 cases and 6,966 controls from the Ovarian Cancer Association Consortium. Given that most of the top 20 SNPs from pooling were validated in the same samples by individual genotyping, the lack of replication is likely to be due to the relatively small sample size in our stage 1 GWAS rather than due to problems with the pooling approach. We conclude that there are unlikely to be any moderate or large effects on ovarian cancer risk untagged by less dense arrays. However, our study lacked power to make clear statements on the existence of hitherto untagged small-effect variants.

■ **Keywords:** GWAS, DNA Pooling, Ovarian cancer risk, Nanodrop spectroscopy.

Genome-Wide Association Studies (GWAS) have been an unprecedented success in identifying common alleles with moderate to small effects associated with different diseases and phenotypes. In particular, more than 100 common, low-penetrance loci of different cancers have been uncovered by GWAS (Varghese & Easton, 2010). The discovery of susceptibility loci will provide significant insights to cancer etiology and an improved understanding of the mechanisms of tumor biology. In addition, loci associated with tumor progression after treatment will offer targets for therapeutic intervention, and the risk predictions based on accumulated knowledge of cancer genetics, together with environmental risk factors, will help to identify individuals with an elevated risk of cancer (Fletcher & Houlston, 2010). Although each of the common loci identified through GWAS only account for a small proportion of risk, collectively more than 20% of familial risk of prostate cancer has been explained, and ~7%, ~6%, and ~5% of familial risk of lung, colorectal, and breast cancers, respectively, can now be explained by GWAS results (Varghese & Easton, 2010). These estimates are likely to be conservative, as the effects of causal variants are typically larger than the associations detected through tag single-nucleotide polymorphisms (SNPs); Fletcher & Houlston, 2010).

Globally, ovarian cancer is the seventh leading cause of cancer mortality among woman. Despite its relatively rare incidence, it has the same pattern of familial aggregation as other major cancers. Early twin studies have shown that most of the excess familial risk of ovarian cancer is due to genetic factors rather than shared environmental factors (Lichtenstein et al., 2000). It is well established that although rare mutations in *BRCA1* and *BRCA2*, identified originally by linkage studies, are the most important genetic risk factors in terms of their high penetrance, they do not fully account for the excess ovarian cancer risk seen in families. To date, one large GWAS has been conducted on ovarian cancer susceptibility aiming to identify some of the remaining unexplained familial risk. This GWAS used relatively low-density Illumina 610K and 550K arrays for cases and controls, respectively Song et al., (2009), Bolton et al., (2010), Goode et al., (2010). The confirmed susceptibility loci reaching genome-wide significance level ($p < 5 \times 10^{-8}$) uncovered by this GWAS are at 9p22.2 near *BNC2* (Song et al., 2009), 19p13.11 near *C19orf62* (also known as *MERIT40*) (Bolton et al., 2010), 8q24, and 2q31; two other borderline significant loci at 3q25 and 17q21 were also identified (Goode et al., 2010). In addition, a candidate gene study has implicated the *TERT* locus, which has been found to contain susceptibility SNPs by many other cancer GWAS (Johnatty et al., 2010). All these loci confer small risks (per-allele relative risk less than 1.3), supporting the concept of polygenic architecture underlying ovarian cancer susceptibility. As found in other cancer types, ovarian cancer also shows histological subtype variation. The associations identified so far are stronger for serous tumors than for other histological subtypes. To identify histology-specific risk loci, separate GWAS on different subtypes will be more powerful than a single GWAS, including all subtypes. However, the high cost of GWAS limits the desirability of carrying out studies with individual genotyping for the less common subtypes, such as endometrioid, mucinous, and clear cell ovarian cancers, which may not be well powered.

The GWAS genotyping on pooled DNA has proved to be a time- and cost-effective alternative to conventional GWAS, which individually genotype all the study subjects (Craig et al., 2009; Macgregor et al., 2006, 2008; Norton et al., 2004; Sham et al., 2002; Visscher & Le Hellard, 2003). In this study we conducted a pooled GWAS on serous ovarian cancer risk, followed by validation of the pooled results and genotyping SNPs of interest in an independent large dataset from the Ovarian Cancer Association Consortium.

We used the high-density Illumina 1M-Duo array containing 1.2 million SNPs for our pooled GWAS, because it has a superior coverage with 93% of common SNPs in the CEU population tagged at $r^2 \geq 0.8$. The aim of this study was two-fold: to test the hypothesis that common SNPs with moderate or low risks, which are not well tagged by the lower density arrays (Illumina 550K and 610K arrays used in the previous GWAS), also account for some of the residual ovarian cancer risk; and to determine whether the pooled GWAS can be effectively carried out on DNA quantified by spectrophotometry, as opposed to Picogreen absorption, which we have used previously.

Materials and Methods

Ethics Statement

This study was conducted according to the principles expressed in the Declaration of Helsinki. The study was approved by the human ethics committee of Queensland Institute of Medical Research. All participants provided written informed consent.

Samples

We used samples from the Australian Ovarian Cancer Study (AOCS) for the pooled GWAS. AOCS ascertained ovarian cancer cases through the surgical treatment centers in Australia, and from the Cancer Registries of Queensland, South and Western Australia, New South Wales, and Victoria, while controls were population-based and drawn from the Commonwealth electoral roll (Burkey & Kanetsky, 2009). We selected 342 invasive serous cases and 643 controls for the pooled GWAS. All the study subjects were self-reported White with non-Hispanic origin. Age at diagnosis and interview was recorded for cases and controls, respectively. Detailed clinical information was also available for ovarian cancer patients, including primary site of tumor, stage and grade, and overall survival time. Most of the DNAs had been isolated using salt extraction (Chang et al., 2009), but a subset had been isolated with Qiagen columns, and so these DNAs were kept separate.

DNA concentration was measured before the pools were made by spectrophometry using a Nanodrop, and the samples were adjusted through serial dilution to 48–52 ng/μL. Each DNA sample of 2 μL was combined for each pool, and the final concentration was verified by Nanodrop. The salt-extracted DNAs from 303 invasive serous cases were further divided into tertiles according to their overall survival time. We made a separate pool of 39 DNAs from unselected cases that were isolated with Qiagen columns. The controls were randomly placed in seven pools, each with a size of 90–91 samples. We then matched each of the three salt-extracted case sets with two of the control sets. The smaller Qiagen-extracted DNA pool was matched with the remaining control pool. Thus, we had four comparisons

TABLE 1
Design of Case-Control Pool Comparisons

Pool comparison	Case-control status[a]	N	Mean age ($\pm SD$)	p-value of mean age difference between cases and controls
1	Case (good survival)	101	58.9 (8.6)	0.3038
	Control	182	57.7 (10.4)	
2	Case (medium survival)	101	59.4 (10.5)	0.1223
	Control	182	57.3 (11.2)	
3	Case (poor survival)	101	62.2 (9.5)	**2.5e-4**
	Control	181	57.3 (12.5)	
4	Case (extracted by Qiagen columns)	39	60.5 (9.9)	**0.0288**
	Control	90	56.2 (10.8)	

Note: [a]We stratified the cases in three large pools by overall survival time. A small subset of 39 case DNAs isolated by Qiagen columns was kept together in one pool.

of case versus control pools, where each individual case was matched with approximately two individual controls (Table 1).

We used samples from 12 sites in the Ovarian Cancer Association Consortium (OCAC) in the replication stage (Table 2). In all, we genotyped 13,779 samples, including 6,966 non-Hispanic White controls and 4,651 non-Hispanic White invasive cases, among which 2,245 were of serous histology.

Genotyping and Quality Control

All the DNA pools were genotyped on Illumina Human 1M-Duo arrays using standard protocols. All pools were genotyped in triplicate, with the exception of one control pool, which was genotyped in quadruplicate. A number of quality control (QC) steps described elsewhere (Lu et al., 2010) were also applied here: (1) SNPs must have less than 10% negative intensity values on each pool; (2) The number of working probes for SNP on each pool must be larger than 20; (3) The sum of raw red and green intensity values must be more than 1,200; (4) Minor allele frequency (MAF) in the HapMap CEU samples is over 5%; (5) SNP must not present significant variance difference between case and control pools. A number of additional checks were also applied. (6) The differential amplification parameter of SNP must be between 1/3 and 3. 'Differential amplification' refers to a phenomenon that the alleles at a locus are unequally amplified; in these cases the allele frequency estimates are biased because of the imbalanced raw intensity value. However, the differential amplification cancels out to a good approximation when we assess the allele frequency difference between case and control pools. We discarded SNPs with very extreme differential amplification (<1/3 or >3). This additional check is equivalent to discarding SNPs with estimated allele frequencies that are very different from the reference samples, for example, the HapMap CEU samples used here. (7) SNPs that passed quality control for more than two pool pairs out of four were kept, because in

TABLE 2

Summary of OCAC Samples Used for the Replication Study

		Controls	Cases		
OCAC site	Study name	Controls (non-Hispanic White)	Cases (non-Hispanic White)	Invasive case (non-Hispanic White)[a]	Invasive serous cases (non-Hispanic White)[b]
AUS[c]	Australia Ovarian Cancer Study/Australian Cancer Study	576 (524)	1,276 (1,207)	921	502
BEL	Belgium Ovarian Cancer Study	428 (428)	257 (253)	197	128
GER	German Ovarian Cancer Study	420 (420)	252 (251)	223	105
HJO/HMO	Hannover–Jena and Hannover–Minsk Ovarian Cancer Study	913 (903)	467 (463)	246	70
HAW	Hawaii Ovarian Cancer Study	625 (166)	417 (102)	81	43
HOCS	Helsinki Ovarian Cancer Study	456 (456)	262 (262)	261	130
NTH	Nijmegen Polygene Study & Nijmegen Biomedical Study	599 (598)	305 (300)	297	101
MAL	Danish Malignant Ovarian Tumour Study	1075 (1075)	263 (263)	263	162
NJO	New Jersey Ovarian Cancer Study	189 (173)	200 (177)	176	105
OVA	Ovarian Cancer in Alberta and British Columbia Study	530 (460)	834 (706)	538	291
SEA	UK SEARCH Ovarian Cancer Study	1,231 (1,227)	1,172 (1,160)	972	377
UKO	UK Ovarian Cancer Population Study	542 (536)	490 (476)	476	231
Total		7,584 (6,966)	6,195 (5,620)	4,651	2245

Note: [a]Cases eligible for secondary analysis in the replication.
[b]Cases eligible for primary analysis in the replication.
[c]AOCS cases and controls included in the pools using stage 1 were excluded from analysis in the replication study.

general the more the working pool pairs, the more the reliable results. (8) For the SNPs of interest, the proxies (linkage disequilibrium (LD) $r^2 > 0.7$) must have similar association results as the underlying SNP. We applied stringent quality controls to limit false positive results rising from pooling design. After a whole series of QC steps, 9,14,948 SNPs were retained.

Individual genotyping for 20 SNPs selected from the pooled GWAS was performed using MALDI-TOF spectrophotometric mass determination of allele-specific primer extension products using Sequenom's MassARRAY system and iPLEX technology (Sequenom Inc.). The design of oligonucleotides was carried out according to the guidelines of Sequenom and performed using MassARRAY Assay Design software (version 4.0). Multiplex PCR amplification of amplicons containing SNPs of interest was performed using the Qiagen HotStart Taq Polymerase on a Perkin Elmer GeneAmp 2400 thermal cycler with 5-ng genomic DNA. Primer extension reactions were carried out according to manufacturer's instructions for iPLEX chemistry. Assay data were analyzed using Sequenom TYPER software (Version 3.4). These SNPs passed the following standard QC checks: (1) *p*-value for the Hardy–Weinberg equilibrium (HWE) test ≥0.05 in both cases and controls; (2) call rate > 95%; (3) concordance >98% between duplicate pairs (at least 5% per study site). One SNP (rs12078260) failed the HWE test in controls.

Statistical Methods and Analytic Tools

In the pooled GWAS, the allele frequencies on each locus were estimated from each pool, and then the differences of allele frequencies between each pair of case/control pool were assessed in the association test. Details of pooling data analysis were described elsewhere (Lu et al., 2010). The four sets of association results from each pool pair were then meta-analyzed, where the allele frequency difference between each set of case and control pools was weighted by its inverse variance (binomial variances in case and control pools plus pooling error variances; Macgregor et al., 2006, 2008). A pooling program that incorporates the steps of estimating pooled allele frequency, mean normalization, quality controls, and finally association test taking into account pool-specific errors, has been developed for the pooled GWAS. This program is available on request.

For individually genotyped data, the SNP association was assessed in a logistic regression model implemented in PLINK (Purcell et al., 2007). Assuming a log additive model of inheritance, the per-allele risk was estimated by fitting the number of rare alleles as continuous variable. We did not adjust for age effect in individual genotyping (IG) validation to allow for a direct comparison of pooled and individual genotyped results on the same AOCS samples. However, the age-adjusted results were similar (results not presented). In the replication stage, both age and study sites were adjusted for in the logistic regression model.

Results

In order to reduce heterogeneity, the majority of invasive serous cases from the AOCS included in the pooled GWAS had tumors that originated in the ovary (except for one case, whose tumor appeared to arise in fallopian tubes), and are of high stage (>92% cases with FIGO (International Federation of Gynaecology and Obstetrics) stage III or IV) and grade (>99% cases with grade 2 or 3). Since age at diagnosis for cases is a predictor of overall survival time, age differences were observed in the comparison of cases with poor survival and controls, which were younger than these cases. A nominally significant difference in mean age was also found in the comparison of cases extracted using Qiagen columns and controls (Table 1). After carrying out

TABLE 3

Comparison of Pooled GWAS and Individual Genotyping (IG) Validation Results for 19 SNPs in AOCS Samples

SNP ID	Chr	Coordinate	Alleles[a]	Pooled GWAS			IG validation		
				OR	SE	P_pool[b]	OR	SE	P_IG
rs7562599	2	153711404	[T/C]	0.595	0.122	2.14E-05	0.637	0.153	3.01E-03
rs10792844	11	85658476	[A/C]	0.463	0.190	4.94E-05	0.886	0.176	4.89E-01
rs1573110	10	9135501	[A/G]	1.744	0.141	8.46E-05	2.003	0.162	1.41E-05
rs17759746	2	28247797	[T/C]	0.596	0.135	1.23E-04	0.649	0.158	5.73E-03
rs8043748	16	11753732	[A/G]	0.686	0.098	1.28E-04	0.679	0.097	6.84E-05
rs17353424	8	107892457	[T/C]	0.638	0.117	1.28E-04	0.565	0.267	3.02E-02
rs7974375	12	117070081	[A/C]	1.608	0.125	1.43E-04	1.680	0.128	4.39E-05
rs10818911	9	125854025	[T/G]	0.627	0.125	1.91E-04	0.656	0.150	4.65E-03
rs4887515	15	85233018	[T/C]	0.610	0.133	1.95E-04	0.573	0.165	6.48E-04
Rs1903532	4	179644249	[T/G]	0.648	0.116	1.98E-04	0.751	0.136	3.46E-02
Rs11592097	10	2166806	[A/C]	0.645	0.121	2.85E-04	0.609	0.194	9.90E-03
Rs2798823	14	94490125	[A/G]	0.687	0.108	4.76E-04	0.701	0.130	6.17E-03
rs2086545	11	13164955	[C/G]	1.450	0.109	6.55E-04	1.154	0.130	2.72E-01
rs2499834	1	160236260	[A/C]	0.627	0.139	7.79E-04	0.623	0.175	6.28E-03
rs1053495	10	71544054	[T/C]	0.611	0.148	8.66E-04	0.779	0.188	1.84E-01
rs1566198	18	11249831	[T/G]	1.510	0.126	1.05E-03	1.373	0.128	1.27E-02
rs16135	7	24294445	[A/G]	0.576	0.170	1.16E-03	0.769	0.197	1.82E-01
rs16899823	5	81992978	[T/C]	0.665	0.133	2.20E-03	0.638	0.197	2.16E-02
rs12027970	1	37537850	[A/C]	0.685	0.124	2.31E-03	0.609	0.213	1.89E-02

Note: [a] The first allele was the risk allele of SNP.
[b] The table was sorted by the strength of association found in pooled GWAS (P_pool).

extensive quality control, we tested association for 9,14,948 SNPs on each comparison of case-control pools (for details of pooling designs, see Samples in Materials and Methods section), and meta-analyzed the four sets of genome-wide association results using a standard weighting method in order to maximize statistical power.

We chose 20 SNPs from the pooled GWAS for individual genotyping in the same AOCS samples as a validation of pooled results. These SNPs were among the top-ranked SNPs from the pooled GWAS that had evidence of association with ovarian cancer susceptibility, but none of these reached genome-wide significance. Moreover, these were selected for being in the subset of SNPs not well tagged by Illumina 610K array, as one of our aims was to test the hypothesis that this pooled GWAS using denser SNP arrays could uncover additional risk SNPs not identified by the previous GWAS. These 20 SNPs were successfully genotyped for nearly all the AOCS samples included in the pooled GWAS (971 out of 985 pooled samples), but one SNP failed quality control. Table 3 compares the odds ratios (OR) and *p*-values from the pooled GWAS and IG validation results. Despite slight difference in samples, good concordance was observed in OR estimates, with all risk directions in agreement in both sets of results. For 15/19 SNPs, the putative associations found in the pooled GWAS were clearly validated in IG results. Therefore, by comparing the results from pooled genotyping and individual genotyping on the same set of samples, we showed that GWAS using pooled DNA, quantified by spectrometry, has the potential to estimate allele frequencies accurately and provide an efficient test of association.

In addition, we sought independent replication for these 19 SNPs by individually genotyping a total of 13,779 samples collected from 12 study sites in OCAC (Table 2). Among 4,651 eligible White invasive cases of non-Hispanic origin, the majority (>95%) were classified as having primary tumor in the ovary, as opposed to the fallopian tube or peritoneum. Unlike the AOCS cases included in the pooled GWAS, the OCAC cases on the whole were evenly distributed over all tumor stages and grades (~54% in high stage and ~58% in high grade). Two sets of analyses were performed according to histology: In the primary analysis we restricted to White non-Hispanic cases with the serous subtype, which allowed for direct replication of SNPs found in the stage 1 GWAS on serous ovarian cancer cases; whereas in the secondary analysis we included cases with all histological subtypes to determine whether these SNPs show association with ovarian cancer regardless of histological types. The association results adjusted for age and study site are presented in Table 4. The results showed no replication for any of the 19 SNPs in the analyses restricted to serous cases only (primary analysis), or in the analyses combining all histological subtypes (secondary analysis).

Discussion

To date, one ovarian cancer GWAS has revealed several SNPs associated with susceptibility. None of the identified loci showed large effects (OR: 0.76–1.30 depending on the histological subtype), but the study was well powered to find common alleles with moderate effects (Song et al., 2009). In contrast, our study of the pooled GWAS on serous ovarian cancer susceptibility was under-powered to detect the alleles with moderate effects because of small sample size. In our pooled GWAS, the published risk SNPs, rs3814113 at 9p22.2, rs2072590 at 2q31, and rs2665390 at

TABLE 4

Replication Results of 19 SNP from Pooled GWAS by Individual Genotyping of 13,779 OCAC Samples

SNP	Chr	Coordinate	Alleles[a]	Primary analysis[c]				Secondary analysis[c]			
				N[b]	OR	SE	p	N[b]	OR	SE	p
rs7562599	2	153711404	[T/C]	9,185	1.06	0.05	0.26	11,580	1.03	0.04	0.48
rs10792844	11	85658476	[A/C]	9,188	0.98	0.06	0.76	11,588	1.06	0.05	0.27
rs1573110	10	9135501	[A/G]	9,184	1.00	0.06	0.96	11,581	0.99	0.04	0.88
rs17759746	2	28247797	[T/C]	9,175	1.03	0.06	0.62	11,572	0.99	0.05	0.80
rs8043748	16	11753732	[A/G]	9,178	1.03	0.04	0.40	11,574	1.03	0.03	0.37
rs17353424	8	107892457	[T/C]	9,198	0.89	0.09	0.24	11,599	0.89	0.07	0.11
rs7974375	12	117070081	[A/C]	9,147	0.98	0.05	0.61	11,547	1.00	0.04	0.97
rs10818911	9	125854025	[T/G]	9,194	1.00	0.05	0.96	11,597	0.98	0.04	0.61
rs4887515	15	85233018	[T/C]	9,175	0.98	0.06	0.77	11,566	1.03	0.05	0.50
rs1903532	4	179644249	[T/G]	9,167	1.08	0.05	0.17	11,551	1.01	0.04	0.84
rs11592097	10	2166806	[A/C]	9,152	0.97	0.06	0.60	11,535	0.95	0.05	0.27
rs2798823	14	94490125	[A/G]	9,184	0.95	0.05	0.34	11,583	0.94	0.04	0.08
rs2086545	11	13164955	[C/G]	9,184	1.03	0.05	0.57	11,581	1.06	0.04	0.19
rs2499834	1	160236260	[A/C]	9,152	0.93	0.06	0.20	11,547	0.96	0.04	0.33
rs1053495	10	71544054	[T/C]	9,200	1.01	0.06	0.93	11,600	1.02	0.05	0.72
rs1566198	18	11249831	[T/G]	9,193	1.03	0.05	0.44	11,593	1.04	0.04	0.30
rs16135	7	24294445	[A/G]	9,187	1.04	0.07	0.55	11,589	1.02	0.05	0.76
rs16899823	5	81992978	[T/C]	9,187	1.04	0.06	0.51	11,586	1.07	0.05	0.14
rs12027970	1	37537850	[A/C]	9,184	0.95	0.07	0.44	11,583	0.97	0.05	0.59

Note: [a]Same risk alleles as listed in Table 2.
[b]Number of samples with non-missing genotypes for each SNP.
[c]Primary analysis was restricted to serous cases; secondary analysis including all histological subtypes.

TABLE 5

Pooled GWAS Results on the Published Loci Known to be Associated with Serous Ovarian Cancer Susceptibility

Position	SNP ID	Publication	A1	Reported per-allele OR[a]	Reported p-value	Pooled per-allele OR	Pooled p-value
9p22.2	rs3814113	Song et al., 2009	C	0.77 (0.73–0.81)	4.1e-21	0.86 (0.72–1.05)	0.082
19p13	rs8170	Bolton et al., 2010	C	1.18 (1.12–1.25)	2.7e-09	0.97 (0.77–1.31)	0.175
	rs2363956		G	1.16 (1.11–1.21)	3.8e-11	0.92 (0.78–1.12)	0.60
2q31	rs2072590	Goode et al., 2010	T	1.20 (1.14–1.25)	3.8e-14	1.21 (1.01–1.41)	0.03
3q25	rs2665390		C	1.24 (1.15–1.34)	7.1e-08	1.22 (0.93–1.79)	0.060
8q24	rs10088218		A	0.76 (0.70–0.81)	8.0e-15	1.28 (0.89–1.66)	0.882
17q21	rs9303542		G	1.14 (1.09–1.20)	1.4e-07	NA[b]	NA[b]

Note: [a]For a direct comparison of results, reported per-allele ORs are the results from the published GWAS restricted to serous cases.
[b]rs9303542 was not on Illumina Human 1Mduo array, but rs6504172 is in high linkage disequilibrium with rs9303542 ($r^2 = 0.841$). It had a per-allele OR in the pooled GWAS of 1.10 (0.84–1.56) ($p = 0.49$).

3q25, showed similar ORs and in the same direction as reported previously (Goode et al., 2010; Song et al., 2009), but these reached nominal or borderline significance only (Table 5). The other three SNPs (rs8170 and rs2363956 at 19p13, and rs10088218 at 8q24) identified by the published GWAS (Bolton et al., 2010; Goode et al., 2010) were not significantly associated with risk in our results, and SNP rs9303542 at the 17q21 (Goode et al., 2010) was not on Illumina Human 1M-Duo array (Table 5). We found a similar OR for rs6504172, which is in high linkage disequilibrium with rs9303542 ($r^2 = 0.841$), but this SNP was not significantly associated with risk ($p = 0.49$). We therefore found no support for our hypothesis that additional common SNPs represented on the 1M-Duo arrays contribute to ovarian cancer risk, probably because of insufficient power in stage 1 of this pooled GWAS.

In the pooling design, we divided serous ovarian cancer cases into four case pools according to the overall survival time and/or the method in which the DNAs were isolated. In theory it would be possible to test for association of SNPs with survival time by comparing good, medium, and poor survival pools, but we would have even less power to detect reliable association with survival, so these results are not presented.

Although we were under-powered to locate any common SNPs with weak effects, our study had the potential to identify common SNPs with moderate to large effects on ovarian cancer risk if any. A notable example in cancer genetics is the common variant in *KITLG* with a per allele risk of 2.5 for testicular cancer, which was identified from an initial GWAS in ∼300 cases and ∼900 controls (Kanetsky et al., 2009). This empirical example suggests that although most loci exhibit smaller effect sizes, common SNPs with moderate to large effect do exist in cancer genetics, and therefore it is of interest to test similar hypothesis in different cancer types. GWAS genotyping on pooled DNA does not suffer from substantial power loss compared to a conventional study using individual genotyping. For example, in our pooled GWAS, assuming an additive effect risk allele with 20% frequency that confers a relative risk of 2, power was 80%

even after scaling the original sample size by 10% in order to account for additional variance because of pooling errors (Macgregor et al., 2008), in comparison with 88% power using individual genotyping. An empirical study with examples of successful identification of the known variants, including the eye color locus at *OCA2*/*HERC2* (15q11.2-q12), the age-related macular degeneration locus at *CFH* (1q32), and the locus for Pseudoexfoliation syndrome at *LOXL1* (15q22) clearly showed that common alleles with large effects are not likely to be missed in the pooled GWAS (Craig et al., 2009). Therefore, our results suggest that there are probably not hitherto poorly tagged common SNPs with moderate to large effects still to be identified.

This study also demonstrates that it is not always necessary to measure DNA concentration by Picogreen absorption, prior to making DNA pools. At least for the set of DNAs we used, which were largely isolated by salt-extraction, this study demonstrated highly consistent results between pooled genotyping and individual genotyping. However, it is worth noting that we have previously found that the correlation between the concentrations measured by Nanodrop spectroscopy and Picogreen adsorption is high ($r^2 = 0.5107$) for a related set of 200 DNAs.

It should be noted that lack of replication in this study was not due to problems in the DNA pooling method. Since additional errors, such as pool construction errors and pool measurement errors, could be involved (Sham et al., 2002), we have implemented careful experiments and rigorous analysis to address this concern. Firstly, we performed careful experiments to ensure equal quantity of DNA contributed by individual samples during the formation of the pools; secondly, all the pools were genotyped at least thrice to yield better allele frequency estimates, and we applied stringent quality controls to limit the number of possible false positives; lastly, we accounted for additional variance because of pooling errors in the association tests. We also validated pooling results using individual genotyping. Given that most of the top 20 SNPs from the pooled GWAS were validated in the same samples by individual genotyping, the lack of replication is most likely to be due to relatively small sample size in our stage 1 GWAS rather than due to problems with the pooling approach.

In order to improve power for GWAS using pooled DNA, larger sample sizes and higher density micro-arrays are required. However, to properly accommodate a large sample size, a balance between the statistical power and the accuracy of the allele frequency estimates is needed. A number of empirical studies have investigated the impact of pool size (up to 1,000 samples in the pool) on the accuracy of allele frequency estimate, and usually found no obvious relationship between the pool size and the accuracy of allele frequency estimation (Jawaid & Sham, 2009; Le Hellard et al., 2002; Macgregor, 2007). As indicated in Macgregor et al. (2007), most variation from pooled DNA genotyping is attributable to array error rather than pool construction

error. Therefore, constructing large pools is not likely to yield a great loss of power. An optimal pool design for a limited research budget will be a few large pools, which are then genotyped for multiple times. One major criticism of the pooled GWAS is that there is no information on individual genotypes or linkage disequilibrium information, so it is generally impossible to impute missing genotypes, evaluate haplotypes, or fine map the regions of interest. However, given the cost advantage, more expensive SNP arrays with better coverage can be used to partially compensate for the power loss because of imperfect linkage disequilibrium between genetic markers and causal variants. Furthermore, fine mapping of loci identified by GWAS is usually performed in a stage 2 or 3 of genotyping once the loci have been confirmed in additional samples. Here we have investigated the use of dense Illumina 1M-Duo arrays in locating variants that were poorly tagged by previous arrays. We found that moderate to large effects on ovarian cancer risk are unlikely to exist among the SNPs on this array, but we are not able to make a clear statement about the possible existence of additional SNPs with small effects because of limited study power.

In summary, we have carried out the pooled GWAS on 342 invasive serous cases and 643 controls. The accuracy of estimated odds ratios was then validated by individually genotyping the same subjects that were included in the pool. We showed that pooled genotyping using DNAs quantified by Nanodrop spectroscopy, together with analytical tools for the pooled data, work well in terms of achieving accurate OR estimations and providing reasonable association signals. We therefore propose to use pooled GWAS for less common subtypes of cancer or orphan diseases where research funds are limited. In addition, we have developed an analytical tool for analyzing the pooled GWAS data, which will be available on request.

Acknowledgments

References

Bolton, K. L., Tyrer, J., Song, H., Ramus, S. J., Notaridou, M., Jones, C., & Gayther, S. A. (2010). Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nature Genetics*, *42*, 880–884.

Burkey, A. R., & Kanetsky, P. A. (2009). Development of a novel location-based assessment of sensory symptoms in cancer patients: Preliminary reliability and validity assessment. *Journal of Pain and Symptom Management*, *37*, 848–862.

Chang, Y. M., Newton-Bishop, J. A., Bishop, D. T., Armstrong, B. K., Bataille, V., Bergman, W., Berwick, M., Bracci, P. M., Elwood, J. M., Ernstoff, M. S., Green, A. C., Gruis, N. A., Holly, E. A., Ingvar, C., Kanetsky, P. A., Karagas, M. R., Le Marchand, L., Mackie, R. M., Olsson, H., Østerlind, A., Rebbeck, T. R., Reich, K., Sasieni, P., Siskind, V., Swerdlow, A. J., Titus-Ernstoff, L., Zens, M. S., Ziegler, A., & Barrett, J. H. (2009). A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *International Journal of Cancer*, *124*, 420–428.

Craig, J. E., Hewitt, A. W., McMellon, A. E., Henders, A. K., Ma, L. J., Wallace, L., Sharma, S., Burdon, K. P., Visscher, P. M., Montgomery, G. W., & MacGregor, S. (2009). Rapid inexpensive genome-wide association using pooled whole blood. *Genome Research*, *19*, 2075–2080.

Fletcher, O., & Houlston, R. S. (2010). Architecture of inherited susceptibility to common cancer. *Nature Reviews Cancer*, *10*, 353–361.

Goode, E. L., Chenevix-Trench, G., Song, H., Ramus, S. J., Notaridou, M., Lawrenson, K., & Pharoah, P. D. (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nature Genetics*, *42*, 874–879.

Jawaid, A., & Sham, P. (2009). Impact and quantification of the sources of error in DNA pooling designs. *Annals of Human Genetics*, *73*, 118–124.

Johnatty, S. E., Beesley, J., Chen, X., Macgregor, S., Duffy, D. L., Spurdle, A. B., & Chenevix-Trench, G. (2010). Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility "hot-spot." *PLoS Genetics*, *6*, e1001016.

Kanetsky, P. A., Mitra, N., Vardhanabhuti, S., Li, M., Vaughn, D. J., Letrero, R., Letrero, R., Ciosek, S. L., Doody, D. R., Smith, L. M., Weaver, J., Albano, A., Chen, C., Starr, J. R., Rader, D. J., Godwin, A. K., Reilly, M. P., Hakonarson, H., Schwartz, S. M., & Nathanson, K. L. (2009). Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. *Nature Genetics*, *41*, 811–815.

Le Hellard, S., Ballereau, S. J., Visscher, P. M., Torrance, H. S., Pinson, J., Morris, S. W., Thomson, M. L., Semple, C. A., Muir, W. J., Blackwood, D. H., Porteous, D. J., & Evans, K. L. (2002). SNP genotyping on pooled DNAs: Comparison of genotyping technologies and a semi-automated method for data storage and analysis. *Nucleic Acids Research*, *30*, e74.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., & Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer analyses of cohorts of twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, *343*, 78–85.

Lu, Y., Dimasi, D. P., Hysi, P. G., Hewitt, A. W., Burdon, K. P., Toh, T., Ruddle, J. B., Li, Y. J., Mitchell, P., Healey, P. R., Montgomery, G. W., Hansell, N., Spector, T. D., Martin, N. G., Young, T. L., Hammond, C. J., Macgregor, S., Craig, J. E., & Mackey, D. A. (2010). Common genetic variants near the Brittle Cornea Syndrome locus ZNF469 influence the blinding disease risk factor central corneal thickness. *PLoS Genetics*, *6*, e1000947.

Macgregor, S. (2007). Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *European Journal of Human Genetics*, *15*, 501–504.

Macgregor, S., Visscher, P. M., & Montgomery, G. (2006). Analysis of pooled DNA samples on high density arrays without prior knowledge of differential hybridization rates. *Nucleic Acids Research*, *34*, e55.

Macgregor, S., Zhao, Z. Z., Henders, A., Martin, N. G., Montgomery, G. W., & Visscher, P. M. (2008). Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Research*, *36*, e35.

Norton, N., Williams, N. M., O'Donovan, M. C., & Owen, M. J. (2004). DNA pooling as a tool for large-scale association studies in complex traits. *Annals of Medicine*, *36*, 146–152.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575.

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA pooling: A tool for large-scale association studies. *Nature Reviews Genetics*, *3*, 862–871.

Song, H., Ramus, S. J., Tyrer, J., Bolton, K. L., Gentry-Maharaj, A., Wozniak, E., & Gayther, S. A. (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nature Genetics*, *41*, 996–1000.

Varghese, J. S., & Easton, D. F. (2010). Genome-wide association studies in common cancers – what have we learnt? *Current Opinion in Genetics and Development*, *20*, 201–209.

Visscher, P. M., & Le Hellard, S. (2003). Simple method to analyze SNP-based association studies using DNA pools. *Genetic Epidemiology*, *24*, 291–296.