Exome genotyping arrays to identify rare and low frequency variants associated with epithelial ovarian cancer risk

Jennifer B. Permuth[1#], Ailith Pirie[2#], Y. Ann Chen[3], Hui-Yi Lin[3], Brett M Reid[1], Zhihua Chen[3], Alvaro Monteiro[1], Joe Dennis[2], Gustavo Mendoza-Fandino[1], AOCS Study Group[4], Australian Cancer Study (Ovarian Cancer)[4], Hoda Anton-Culver[5], Elisa V. Bandera[6], Maria Bisogna[7], Louise Brinton[8], Angela Brooks-Wilson[9,10], Michael E. Carney[11], Georgia Chenevix-Trench[4], Linda S. Cook[12], Daniel W. Cramer[13], Julie M. Cunningham[14], Cezary Cybulski[15], Aimee A. D'Aloisio[16], Jennifer Anne Doherty[17], Madalene Earp[18], Robert P. Edwards[19,20], Brooke L. Fridley[21], Simon A. Gayther[22], Aleksandra Gentry-Maharaj[23], Marc T. Goodman[24,25], Jacek Gronwald[15], Estrid Hogdall[26], Edwin S. Iversen[27], Anna Jakubowska[15], Allan Jensen[28], Beth Y. Karlan[29], Linda E. Kelemen[30], Suzanne K. Kjaer[28,31], Peter Kraft[32], Nhu D. Le[33], Douglas A. Levine[7], Jolanta Lissowska[34] Jan Lubinski[15], Keitaro Matsuo[35], Usha Menon[23], Rosemary Modugno[19,20,36], Kirsten B. Moysich[37], Toru Nakanishi[38], Roberta B. Ness[39], Sara Olson[40], Irene Orlow[40], Celeste L. Pearce[22,41], Tanja Pejovic[42,43], Elizabeth M. Poole[44,45], Susan J. Ramus[22], Mary Anne Rossing[46,47], Dale P. Sandler[48], Xiao-Ou Shu[49], Honglin Song[50], Jack A. Taylor[48], Soo-Hwang Teo[51,52], Kathryn L. Terry[13], Pamela J. Thompson[29], Shelley S. Tworoger[44,45], Penelope M. Webb[53], Nicolas Wentzensen[54], Lynne R. Wilkens[55], Stacey Winham[56], Yin-Ling Woo[57], Anna H. Wu[22], Hannah Yang[54], Wei Zheng[58], Argyrios Ziogas[59], Catherine M. Phelan[1], Joellen M. Schildkraut[60,61], Andrew Berchuck[62], Ellen L. Goode[18], Paul DP Pharoah[2,50], Thomas A. Sellers[1*] on behalf of the Ovarian Cancer Association Consortium

[#]shared co-first authorship

Corresponding Author:
Thomas A. Sellers, PhD, MPH
Director, Moffitt Cancer Center
Executive Vice President
12902 Magnolia Drive
Tampa, FL 33612
813-745-1315

1. Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA.
2. Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK.
3. Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA.
4. Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Queensland, AUS
5. Department of Epidemiology, Director of Genetic Epidemiology Research Institute, UCI School of Medicine, University of California Irvine, Irvine, CA, USA.
6. Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA
7. Gynecology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
8. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.
9. Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver BC, Canada.
10. Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, BC, Canada.
11. Department of Obstetrics and Gynecology, John A Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, USA
12. Division of Epidemiology and Biostatistics, Department of Internal Medicine, University of New Mexico, Albuquerque, New Mexico, USA.
13. Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Boston, Massachusetts, USA.
14. Department of Laboratory of Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.
15. International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.
16. Social and Scientific Systems, Inc., Durham, NC, USA.
17. Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, NY, USA.
18. Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA.
19. Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.
20. Ovarian Cancer Center of Excellence, Womens Cancer Research Program, Magee-Womens Research Institute & University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA.
21. Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS, USA.
22. Department of Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, California, USA.
23. Women's Cancer, Institute for Women's Health, University College London, London, United Kingdom.
24. Cancer Prevention and Control, Samuel Oshin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
25. Community and Population Health Research Institute, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
26. Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark and Molecular Unit, Department of Pathology, Herlev Hospital,

University of Copenhagen, Copenhagen, Denmark.
27. Department of Statistical Science, Duke University, Durham, NC, USA.
28. Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark.
29. Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California.
30. Department of Public Health Sciences, Medical University of South Carolina College of Medicine, Charleston, SC, USA.
31. Department of Gynaecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.
32. Program in Genetic Epidemiology and Statistical Genetics, Harvard TH Chan School of Public Health, Boston, MA, USA.
33. Cancer Control Research, BC Cancer Agency, Vancouver, BC, Canada.
34. Department of Cancer Epidemiology and Prevention, The Maria Sklodowska-Curie Memorial Cancer Center, Warsaw, Poland.
35. Division of Molecular Medicine, Aichi Cancer Center Research Institute, Nagoya, Aichi, Japan.
36. Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA.
37. Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA.
38. Department of Gynecologic Oncology, Aichi Cancer Center Central Hospital, Nagoya, Aichi, Japan.
39. The University of Texas School of Public Health, Houston, TX, USA.
40. Memorial Sloan Kettering Cancer Center, Department of Epidemiology and Biostatistics, New York, NY, USA.
41. Department of Epidemology,University of Michigan School of Public Health, Ann Arbor, MI, USA.
42. Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA.
43. Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon, USA.
44. Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
45. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.
46. Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.
47. Department of Epidemiology, University of Washington, Seattle, WA, USA.
48. Epidemiology Branch, Division of Intramural Research, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA.
49. Vanderbilt University School of Medicine, Nashville, TN, USA.
50. Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK.
51. Cancer Research Malaysia, Subang Jaya, Malaysia.
52. University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia.
53. Population Health Department, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia.
54. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda MD, USA.
55. Cancer Epidemiology Program, University of Hawaii Cancer Center, Hawaii, USA.

56. Department of Health Science Research, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA.
57. Department of Obstetrics and Gynaecology, University Malaya Medical Centre, University Malaya, Kuala Lumpur, Malaysia.
58. Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA.
59. Department of Epidemiology, University of California Irvine, Irvine, CA, USA.
60. Department of Community and Family Medicine, Duke University Medical Center, Durham, NC, USA.
61. Cancer Control and Population Sciences, Duke Cancer Institute, Durham, NC, USA.
62. Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina, USA.

Exome genotyping arrays to identify rare and low frequency variants associated with epithelial ovarian cancer risk

ABSTRACT

Rare and low frequency variants are not well covered in most germline genotyping arrays and are understudied in relation to epithelial ovarian cancer (EOC) risk. To address this gap, we used genotyping arrays targeting rarer protein-coding variation in 8,165 EOC cases and 11,619 controls from the international Ovarian Cancer Association Consortium (OCAC). Pooled association analyses were conducted at the variant and gene level for 98,543 variants directly genotyped through two exome genotyping projects. Only common variants that represent or are in strong linkage disequilibrium (LD) with previously-identified signals at established loci reached traditional thresholds for exome-wide significance ($P<5.0x10^{-7}$). One of the most significant signals ($P_{all\ histologies}=1.01x10^{-13}$;$P_{serous}=3.54x10^{-14}$) occurred at 3q25.31 for rs62273959, a missense variant mapping to the *LEKR1* gene that is in LD ($r^2=0.90$) with a previously identified 'best hit' (rs7651446) mapping to an intron of *TIPARP*. Suggestive associations ($5.0x10^{-5}>P>=5.0\ x10^{-7}$) were detected for rare and low-frequency variants at 16 novel loci. Four rare missense variants were identified (*ACTBL2* rs73757391 (5q11.2), *BTD* rs200337373 (3p25.1), *KRT13* rs150321809 (17q21.2) and *MC2R* rs104894658 (18p11.21)), but only *MC2R* rs104894668 had a large effect size (OR=9.66). Genes most strongly associated with EOC risk included *ACTBL2* ($P_{AML}=3.23\ x\ 10^{-5}$; $P_{SKAT-o}=9.23x10^{-4}$) and *KRT13* ($P_{AML}=1.67\ x\ 10^{-4}$; $P_{SKAT-o}=1.07x10^{-5}$), reaffirming variant-level analysis. In summary, this large study identified several rare and low-frequency variants and genes that may contribute to EOC susceptibility, albeit with possible small effects. Future studies that integrate epidemiology, sequencing, and functional assays are needed to further unravel the unexplained heritability and biology of this disease.

INTRODUCTION

Epithelial ovarian cancer (EOC) has a strong heritable component, with an estimated three-fold increased

risk among women with a first-degree relative having the disease (1). The excess familial risk that is not attributed to high penetrance mutations in genes such as *BRCA1* and *BRCA2* may be due to a combination of common and rare alleles that confer low- to moderate penetrance(2, 3). Genome-wide association studies (GWAS) of EOC that have been conducted using most of the samples included in the current investigation have identified common variants at approximately 22 loci that collectively account for 4% of the estimated heritability (4-13). Few data exist regarding the contribution of rare (minor allele frequency (MAF) <0.5%) and low frequency (MAF 0.5-5%) protein-coding variants to EOC risk. This reflects the fact that protein- coding variants have not been targeted by conventional GWAS(14) despite prediction that their effects could be substantial(15) and imputation is known to be challenging for rare variants(16).

Following GWAS arrays of the mid-2000s, exome-based arrays were developed in 2012. The Affymetrix Axiom ® Exome Genotyping Array and the Illumina HumanExome Beadchip each contain >245,000 putative functional coding variants and other categories of variants selected from 16 exome sequencing initiatives that included approximately 12,000 individuals of diverse ethnic backgrounds and a range of diseases(17) (**Supplementary Table 1**). Variants were included as 'fixed' content on the arrays if they occurred at least three times and were seen in two or more of the 16 studies (17). Here, we report the first large-scale genetic association study of uncommon exome-wide variants and EOC risk among nearly 20,000 women (**Supplementary Table 2**).

RESULTS

Of the 98,299 polymorphic variants successfully genotyped as part of EOC case-control set 1 and set 2 (7,308 cases and 10,773 controls; **Supplementary Figure 1**), most (68%) were rare (MAF < 0.5%), many (20%) were common (n=19,565, MAF >5%), and 12% (n=12,175) were low frequency (MAF between 0.5% and 5%). The majority of these variants were non-synonymous (87%) with 81% missense, 1% nonsense,

4% located in splice sites, and <1% resulting in a frame shift.

Single variant associations

The quantile-quantile (Q-Q) plot of the distribution of test statistics for the comparison of genotype frequencies in cases and controls showed slight inflation in the median test statistics of the likelihood ratio tests (λ=1.15; **Supplementary Figure 2**). This slight inflation may be explained by properties of the likelihood ratio test which make it sensitive to rare variants(18). No rare or low-frequency variants were statistically significantly associated with EOC risk ($P<5.0x10^{-7}$); only common variants that represent or are in strong linkage disequilibrium (LD) with previously-identified signals at established loci (2q31.1, 3q25.31, 8q24.21, 9p22.2, 17q12, 17q21.3, and 19p13.1) reached traditional thresholds for exome-wide significance ($P<5.0x10^{-7}$) (**Figures 1A and 1B, Supplementary Table 3**). Briefly, the most statistically significant association was observed at 9p22.2 for a previously identified intronic variant near the *BNC2* (*basonuclin2*) gene(4), rs38114113, with an odds ratio (OR), 95% confidence interval (CI), and P-value of 0.78 (0.75-0.82) ($P=2.96 x10^{-24}$) and 0.75 (0.72-0.79) ($P=3.32x10^{-28}$) among all histologies and serous histology, respectively. rs38114113 is correlated (r^2=0.57-0.95) with two other detected SNPs ($P=10^{-18}$) near *BNC2*. The full genome-wide set of summary association statistics are given in Supplementary Table 4.

The next most significant signal was at 3q25.31, with rs62273959 $P=1.01x10^{-13}$ and $P=3.54x10^{-14}$ in all histologies (OR=1.41) and serous only (OR=1.45) analyses. rs62273959 is a missense variant mapping to the *LEKR1* (*leucine, glutamate and lysine rich 1*) gene which is in LD (r^2=0.90) with a previously identified 'best hit', rs7651446(12) that is located in an intron of *TIPARP (TCDD-inducible poly(ADP- ribose) polymerase)*. Imputation of the region (see **Supplementary Methods**) identified rs78561123 (T>C) ($P=2.97x10^{-15}$), a novel top-ranking variant that maps within 0.5kb of the 3'UTR for *LINC00886* (*long intergenic non-protein coding RNA 886*) and is in strong LD with rs7651446 (r^2=0.97) and rs62273959

(r^2=0.93) (**Supplementary Figure 3A**). The minor alleles of these three variants are located within the same haplotype associated with an increased risk among all histologies (OR (95% CI=1.41 (1.29-1.55), P=3.14x 10^{-13}). A fixed-effect meta-analysis of our study with an imputed dataset from the COGS genotyping initiative (19) also revealed stronger associations for variants located near *TIPARP* and *LINC00886* (**Supplementary Figure 3B**). The combined analysis of set 1 and set 2 also confirmed the existence of known common EOC susceptibility alleles or their proxies at 17q12, 8q24.21, 17q21.3, and 19p13.1 ($P<5.0x10^{-7}$) (**Supplementary Table 3**).

Associations for variants reaching a less stringent threshold ($5.0x10^{-5}>P>=5.0\ x10^{-7}$) were detected among all histologies, serous histology, or endometrioid histology at 16 novel loci (1p36.33, 2p22.1, 3p25.1, 3p14.2, 5q11.2, 6p22.1, 6p21.33, 6q25.2, 6p12.1, 8q21.13, 11q13.1, 15q12, 16q22.3, 17q.21.2, 18p11.21, and 22q11.2) (**Table 1;Figures 2A and 2B**). Of the novel variants that were identified, most were common and four were rare ($MAF_{controls}<0.003$). The four rare missense variants (map to *actin, beta-like 2, ACTBL2* (5q11.2), *biotinidase, BTD* (3p25.1), *keratin 13 type I, KRT13* (17q21.2), and *melanocortin 2 receptor, MC2R* (18p11.21). Visual inspection of cluster plots for all four rare variants underscored that the variant calling was good. Regional association plots for each of these rare variants reveal that they do not appear to be strongly correlated with other genotyped variants (**Figure 3**). The identified rare variants mapping to *ACTBL2, BTD,* and *MC2R* are predicted to be damaging per Polyphen-2 (**Table 1**). Due to low heterozygous genotype counts, it was not possible to estimate ORs for variants at *ACTBL2* and *BTD*. For rs150321809 in *KRT13* and rs104894658 in *MC2R*, the magnitudes of association were relatively high, with ORs of 2.24 and 9.66, respectively, among all histologies. Analysis of 883 invasive endometrioid cancers identified three common variants at $P<5.0x10^{-5}$ (**Figure 2C; Table 1**). We recently described the contribution of deleterious coding variants in seven putative EOC susceptibility genes *(BRIP1, BARD1, PALB2, NBN, RAD51B, RAD51C,* and *RAD51D)* to EOC risk (23, 24). The pooled exome dataset was used

to examine associations for 68 variants that reside in these 7 genes. None of these variants reached levels of statistical significance ($P<5.0x10^{-5}$) in overall, serous, or endometrioid specific analyses. The most significant rare variant in overall and serous analyses was *BRIP1 rs4988345 (*MAF=0.0047; P=0.022 and P=0.024, respectively), whereas *PALB2* rs57605939 (MAF=0.0002) was the variant most significantly associated with endometrioid cancer risk (P=0.007). Similarly, we followed up on an exome sequencing study of 429 serous EOC cases and 557 controls by Kanchi et al.(25) in which rare truncation and missense variants were detected in known EOC susceptibility genes including *BRCA1, BRCA2, CHEK2,* and *PALB2* and in genes not previously associated with EOC susceptibility such as *NF1* and *CDKN2B*. Only four of the rare truncation or missense variants they(25) identified were represented on either of the genotyping arrays utilized in the current investigation. Applying a threshold of P<0.05 for these four variants (BRCA1_772, CLTC_1498, ERCC2_635, and ITK_448) only BRCA1_772 (p.Val772Ala; rs80357467, MAF= 0.00033) was associated with overall EOC susceptibility in our pooled analysis (OR (95%CI): =4.64 (1.22-17.7)), with P=0.014 (serous OR=3.79, P=0.043). This variant is classified as non-pathogenic for the purposes of clinical management but may have a mild to moderate impact on risk(26). Thus, previously- detected rare variants were not strongly associated with EOC susceptibility in our larger dataset.

We also evaluated association results for the 80,178 set 1 variants (N=5,431 case and 5,639 controls) that were not in the pooled dataset (**Supplementary Figure 1**). Results for the most statistically significant ($P<5.0x10^{-5}$) set 1 variants are displayed in **Supplementary Figures 4A-C** and are summarized in **Supplementary Table 5**. Of six set 1 variants that were detected at the $P<5.0x10^{-5}$ threshold of statistical significance, the most statistically significant association was again with a common variant at a known locus rs62273902 (MAF=0.06) in the 5' untranslated region of *LEKR1* showing an increased EOC risk among all histologies (OR=1.42, $P=1.91x10^{-9}$) and serous histology (OR=1.46, $P=9.48x10^{-10}$) that is strongly correlated ($r^2=0.95$) with rs62273959, a variant identified in the pooled analysis (**Supplementary Table 3**).

Of the remaining five variants, one is rare and two have low frequencies. Rare variant rs115783655 (T>C) (MAF=0.005) maps to an intron in *IZUMO4* (*IZUMO family member 4*) and was associated with endometrioid cancer risk ($P=3.32 \times 10^{-5}$) while low frequency missense variants chr19:38572993 (A>G, MAF=0.02) in *SIPA1L3* (*signal-induced proliferation- associated 1 like 3*) and rs148738146 (T>C, MAF=0.01) in *PLA2G12A* (*phospholipase A2, group XIIA)* were associated with decreased risks of EOC in all histologies and serous histology analyses (**Supplementary Table 5**).

Eight of 9,600 indels assessed in set 1 only reached a threshold of $P<9.0\times10^{-4}$, and only one of these (rs147613544 at 8p21.3) is rare (MAF=0.0009) and was associated with a decreased risk for EOC (OR =0.16, $P=5.0\times10^{-4}$). Set 1 also assessed 146 variants in the mitochondrial genome; only one rare (MAF=0.003) non-synonymous variant c.6480G>A (p.Val193Ile) located within *cytochrome c oxidase subunit 1* (COI) was strongly associated with decreased EOC susceptibility among all histologies (OR=0.54, P=0.0009) and serous histology (OR=0.24, P=0.0008). G6480A has been associated with an increased risk for prostate cancer in African Americans (27).

Gene-level associations

In combined analysis of Affymetrix- and Illumina-based data, thirteen genes had P-values less than 5×10^{-4} for an association with EOC susceptibility overall based on the RAML test(28) (**Figure 4**). Consistency was observed when comparing gene-level findings from RAML to those based on the SKAT-O(29) tests (**Table 2**). The genes that were most strongly associated with EOC risk using RAML included actin, beta-like 2, *ACTBL2* ($P_{AML}=3.23 \times 10^{-5}$; $P_{SKAT\text{-}o}=9.23\times10^{-4}$) and *keratin 13, KRT13* ($P_{AML}=1.67 \times 10^{-4}$; $P_{SKAT\text{-}o}=1.07\times10^{-5}$); these genes contained individual variants (rs73757391 and rs150321809) associated with EOC risk ($5.0\times10^{-5}>P>=5.0\times10^{-7}$) highlighted in **Table 1**. Details regarding the set of rare variants that contributed to gene-level findings for *ACTBL2* and *KRT13* are summarized in **Supplementary Table 6.**

ACTBL2 and *KRT* were also statistically significant in the serous-only analysis after multiple correction testing using a FDR threshold of 15%(30). Of the genes featured in **Table 2**, *MC2R* also contained individual rare variants associated with EOC risk ($5.0x10^{-5}$>P>=$5.0x10^{-7}$) (**Table 1**). When comparing primary high-grade serous EOC tumors and normal fallopian tube tissues in the TCGA dataset, two of the aforementioned genes were differentially expressed: *KRT13* was overexpressed in tumor versus normal tissue (p=0.034) while *MC2R* was under-expressed (p=0.004), though neither finding was significant after adjustment for multiple comparisons.

Gene-level results for the 15,042 genes encompassed by set 1 only variants did not highlight *ACTBL2* , *KRT13, or MC2R.* Rather, *leukocyte receptor tyrosine kinase (LTK)* (P=$2.22x10^{-5}$), *ATPase NA+/K+ transporting alpha 3 polypeptide* (*ATP1A3*) (P=$8.33x10^{-5}$), and *son of sevenless homolog 2* (*SOS2*) (P=$4.55x10^{-5}$) were identified as the most strongly associated genes among all histologies, serous histology, and endometrioid histology, respectively (**Supplementary Table 7; Supplementary Figure 5**). Collectively, all genotyped uncommon variants (MAF<0.05%) explained 4.7% of the phenotypic variation in our subjects (54). Only 2% (.11/4.7) of this variation can be attributed to variants with P< $5.0x10^{-5}$ (**Supplemental Methods**).

DISCUSSION

We report an EOC risk association analysis of 98,299 variants enriched for rare and low frequency protein-coding changes among nearly 20,000 women using commercially available genotyping arrays. Assuming a disease prevalence of 1.4%, our sample size was adequately powered (~89%) to detect associations with low frequency variants included on the exome arrays and moderate effect sizes (OR>1.35) should they exist, but we did not identify any novel uncommon variants at exome-wide levels of statistical significance (P<$5.0x\ 10^{-7}$). Instead, association with common variants (MAF>5%) at known EOC

loci (2q31.1, 3q25.31, 8q24.21, 9p22.2, 17q12, 17q21.3, and 19p13.1)(4-6, 10, 19) were identified; most of these variants were (or were in strong LD with) the previously reported top-ranking variant at the locus. Importantly, sixteen novel loci with low-frequency or rare variants at $P<5.0 \times 10^{-5}$ were detected. Four rare variants were identified (*ACTBL2* rs73757391 (5q11.2), *BTD* rs200337373 (3p25.1), *KRT13* rs150321809 (17q21.2), and *MC2R* rs104894658 (18p11.21)), and gene-level analyses revealed statistically significant associations with variation in three of these genes. These results are consistent with the known landscape of common genetic variation in EOC risk and the utility of multi-marker testing for rare variation. They suggest that the effect sizes of rare coding variants with MAF that are in the range included on the exome arrays may be less than 1.35, requiring larger sample sizes and the use of a family-based approach for their discovery. Indeed, recent simulation studies suggest that sample sizes of 60,000-100,000 will be needed to detect small effect sizes for rare variants (MAF<0.5%) when using exome genotyping arrays(31).

Among the four rare variants that were identified, rs73757391, rs200337373, and rs104894658 are non-synonymous and predicted to be damaging. Moreover, according to the ClinVar database(26), *BTD* rs200337373 (G>A) and *MC2R* rs104894658 (C>A) are reported to be pathogenic for biotinidase deficiency and adrenocorticotropic hormone (ACTH) resistance observed in familial glucocorticoid deficiency(32), respectively. Importantly, gene-level analysis using different methods highlighted *ACTBL2, KRT13*, and *MC2R* as being strongly associated with EOC risk overall and serous disease. Whereas *ACTBL2* is a cytoskeletal protein abundantly expressed in vascular smooth muscle cells(33) that has no reported link to cancer, *BTD* (biotinidase) is a putative biomarker of breast cancer(34), papillary thyroid cancer aggressiveness(35), and lymph node involvement in patients with early stage cervical cancer(36). *In vivo* a deficiency of biotinidase affects the expression of central-carbon metabolism genes(37), a pathway important in the development and progression of EOC(38, 39). *KRT13* encodes a cytoskeletal protein downregulated in an estrogen receptor (ER) positive ovarian cancer cell line(40) and contributes to breast cancer growth and metastasis through its interaction with estradiol and the selective estrogen receptor modulators, tamoxifen and raloxifene(41). *MCR2*

belongs to a family of melanocortin receptors involved in the regulation of food intake, inflammation, skin pigmentation, sexual function, and steroidogenesis, in part by binding to adrenocorticotropic hormone (ACTH)(42, 43). ACTH-producing ovarian tumors have been reported (44-46), but this has been in the context of Cushing's syndrome and non-epithelial ovarian cancers. Taken together, there is biological plausibility to explain some but not all of the current association results.

Independent replication of novel rare variant associations is important but challenging because of the lack of appropriate replication panels. The large COGS EOC meta-GWAS(19) with imputation to Phase I 1000 genome project data was completed after the onset of this study. As a form of replication, we attempted to interrogate this dataset(19) for the most strongly associated novel rare variants. Unfortunately, rare variants and their proxies were not represented in the imputed dataset, precluding the possibility of replication and the opportunity to evaluate associations between germline genotype and gene expression via expression quantitative trait locus analysis. Furthermore, our attempt to replicate associations with rare variants identified in studies of EOC that were much smaller than ours (23-25) did not yield statistically significant findings.

The limited evidence for novel rare or low frequency coding variants at exome-wide levels of significance is consistent with studies of other complex diseases (myocardial infarction(51), Alzheimer's disease(52), and insulin processing and secretion(53)) that used these exome genotyping arrays. Published investigations of exome genotyping array data are limited for other cancers, precluding comparison of findings across cancer types. The limited evidence for rare or low-frequency coding variants may be expected when using rare variant chips for cohorts or diseases for which they were not originally designed. Integration of sequencing data in very large sample sizes may be an effective strategy for discovering additional rare EOC alleles in the future since the arrays do not provide complete coverage of all functional variants at each locus and the accuracy of imputation for rare variants is suboptimal. Even with such limitations, the current study suggests that rare coding variants with large effects may exist, although they did not account for a significant fraction of EOC heritability within our data. In total, rare variants accounted for 4.7% of the

phenotypic variation in our subjects (54) with only 2% (.11/4.7) of this variation from variants with P< $5.0x10^{-5}$ (**Supplemental Methods**). The remaining 98% of variance attributable to rare variants in this study could be due to small effect sizes that did not reach statistical significance. In the absence of opportunities to significantly increase sample sizes, future studies should rely on closer integration of epidemiology and laboratory assays of functional effects to further unravel the etiology of this disease.

MATERIALS AND METHODS

Study Population and Genotyping

Study participants came from 27 independent studies in the international Ovarian Cancer Consortium (OCAC)(55) (**Supplementary Table 2**). In brief, cases were women with pathologically-confirmed primary invasive EOC, fallopian tube cancer, or peritoneal cancer, and controls were women without EOC, with at least one ovary intact, and for most studies were frequency-matched to controls on age group and self-reported race. Specimens and data were collected according to protocols approved by local institutional review and ethics boards. Germline DNA samples from 19 studies (Set 1, 7,060 EOC cases and 6,712 controls) were genotyped on the Affymetrix Axiom Exome Genotyping Array at the Affymetrix Service Lab (Santa Clara, CA, USA), and those from eight studies (Set 2, 2,109 cases, 5,646 controls) were genotyped on the Illumina HumanExome Beadchip at the Strangeways Research Lab (University of Cambridge, United Kingdom).

Genotyping Quality Control (QC)

Set 1 genotyping was performed in batches grouped according to sample type (genomic blood, genomic saliva, whole genome-amplified (WGA) blood, WGA saliva). Affymetrix Genotyping Console™ Software was used for automated allele calling for each batch, followed by initial sample and variant QC performed per protocol (http://media.affymetrix.com/support/downloads/manuals/axiom). Since significant batch effects

were observed, intensity data from the genomic samples were combined into a single batch to enable the automated clustering algorithm to more accurately detect rare variants(56). WGA samples were not recalled as one batch because of known chemistry differences between the component batches (personal communication, Affymetrix, Inc). Four hundred thirty-seven samples were genotyped in duplicate and were identified with 99.8% concordance. As shown in **Supplementary Figure 1**, of 13,772 unique samples that were genotyped, 454 (3.3%) were excluded because they failed Affymetrix QC metrics (<97% call rate or dish QC < 0.82) and an additional 545 samples were excluded because of ambiguous gender, replicate discordance, sample relatedness, or failure to meet eligibility criteria for the primary analysis. Of 302,461 variants on the Affymetrix array, 123,934 variants (41%) were excluded for QC reasons which mostly included failed Affymetrix cluster QC, monomorphism, deviation from Hardy Weinberg Equilibrium (HWE) $P<10^{-7}$ in controls, or discordant B allele frequencies between the genomic and WGA samples. A total of 12,773 samples (6,288 case and 6,485 controls) and 178,527 variants genotyped on the Affymetrix platform passed QC steps. HapMap DNA samples for European (CEU, n=60), African (YRI, n=53) and Asian (JPT+CHB, n=88) populations were also genotyped, and the program LAMP(57) was used to estimate intercontinental ancestry based on the HapMap (release no. 23) genotype frequency data for the European, Asian, and African populations. Subjects with greater than 90% European ancestry were included in analyses (5,431 cases, 5,639 controls) (**Supplementary Figure 1**). Genotype data for set 1 are being released into dbGAP per NIH guidelines.

Set 2 genotyping was performed for 7,612 samples, and genotype calling was carried out according to Best Practice Guidelines(56) using the GenCall(58) module in Illumina's Genome Studio with a default GenCall threshold of 0.15. After initial sample QC, zCall(59) was calibrated to a z-value of 8 and run for all variants. One hundred and forty-three samples were genotyped in duplicate and identified with 92% concordance. Initial sample QC excluded 248 samples with low call rates (<70%). After zCall calibration

and variant recall, we further excluded 221 (3%) samples for reasons including <99% call rate, high or low heterozygosity at a significance level of 10^{-16}, ambiguous gender, relatedness, or genotypes discordant with prior genotypes from the international Collaborative Oncological Gene-Environment Study (iCOGS) genotyping array(19). Genotyping also included HapMap DNA samples (CEU, n=95; YRI, n=82; JPT+CHB, n=93), and ancestry was assessed using the IBS matrix for all samples combined with HapMap samples over the uncorrelated variants. Using this multi-dimensional scaling on a weighted identity by state matrix, non-European samples at a distance of greater than 10% were excluded (n=97) (**Supplementary Figure 1**). Of the 247,870 markers on the Illumina array, we excluded 94,231 variants (38%) for reasons including call rate <95%, poor cluster separation, duplicate probes, monomorphism, and deviation from HWE (**Supplementary Figure 1**). We tested for HWE using a Robertson and Hill test statistic stratified by study(60) and an exact test. Variants that failed both tests were excluded using exclusion thresholds of p-values < 10^{-12} and 10^{-6} for cases and controls, respectively. After all exclusions there were 7,046 European ancestry samples (1,878 cases and 5,168 controls) and 153,639 variants genotyped on the Illumina platform (**Supplementary Figure 1**). Thirty-five samples were genotyped in common as part of set 1 and set 2; the genotype concordance rate was 99.66%.

Thus, a combined total of 18,081 unique subjects of European ancestry (7,308 cases, 10,773 controls) were genotyped and passed sample QC. Of the 18,081 subjects, 5,138 (997 cases and 4,141 controls) were not previously genotyped as part of a previously described EOC GWAS or post-GWAS initiative (4, 5, 19). Of the variants passing QC for each set, 98,543 were present on both platforms and available for pooled analysis (**Supplementary Figure 1**), excluding non-autosomal variants (n=1,983) and those with discordant B allele frequencies between the two sets (n=32). On this combined dataset, we carried out principal components analysis (PCA) to examine the sub-European population structure using a linkage-disequilibrium-pruned set of 27,335 autosomal markers with MAF>1% and HWE P value > 10^{-7}. We inspected the first 10 principal

components (PC) for evidence of population stratification in the pooled samples.

Single Variant Analysis

Each variant was tested for a per allele association with EOC risk using a likelihood ratio test comparing the deviance (-2 × log-likelihood) of two generalized linear models with and without the variant. Models were adjusted for set (1 versus 2) and the first five PCs representing sub-European ancestry. When adjusted for study alone, there was an inflation of the test statistics ($\lambda = 1.20$, $\lambda_{1000} = 1.024$) which was reduced to $\lambda = 1.15$, ($\lambda_{1000} = 1.018$) after adjustment for five principal components (61). Visual inspection of intensity cluster plots resulted in the elimination of 244 variants with poor differentiation between heterozygote and homozygote calls (**Supplementary Figure 1**). Subgroup analysis was conducted for the two most common histologic subtypes: serous and endometrioid. Using a stringent Bonferroni correction for 98,299 tests, we considered variants with $P<5.0x10^{-7}$ to be statistically significant. Because of the greater number of variants and samples in set 1, we also explored associations for 80,178 variants from set 1 that passed visual cluster inspection and were not included in the Illumina array; for these analyses, we adjusted for the first 3 PCs.

Gene-level Analysis

Given the emphasis of each array on exomic coverage(62), gene-level tests were also conducted, mapping variants within 50 kb to genes based on Genome Build 37 coordinates and gene annotation that was curated by Affymetrix from UCSC Genome Bowser data tables. In total, 71,044 variants mapped to 15,118 genes of which 12,123 genes contained more than one variant and were evaluated for the pooled gene-level analysis. Two methods were used for gene-level analyses because of some similarity in assumptions and the ability to include covariates in the underlying regression model: a) the rare admixture maximum likelihood test (RAML)(63), which makes no assumptions about the proportion of variants that are associated with

the phenotype of interest or the magnitude and direction of their effect and b) the Sequence Kernel Association Test -Optimal unified test (SKAT-O)(64), a score-based variance-component test that is powerful when the direction of association for variants can be increased or decreased. Both methods considered only rare variants (MAF <1%) and were not weighted based on MAF. False discovery rate (FDR) is used to adjust for multiple comparisons and FDR of 15% is used to declare significance. Similarly, we also conducted gene-level analysis with the larger set of variants in set 1 which totaled 15,042 genes and 128,992 variants. For genes that were most strongly associated with EOC susceptibility, we mined publicly available gene expression data from the Cancer Genome Atlas Project (TCGA)(65) and compared gene expression between 568 high-grade serous ovarian tumors and 10 normal fallopian tube tissues according to previously described methods(9).

ACKNOWLEDGEMENTS

The authors thank the individuals who participated in this research and all the researchers, clinicians, and staff who made possible the many studies contributing to this work.

Funding for this study was supported by the National Institute of Health and the Genetic Associations and Mechanisms in Oncology (GAME-ON), a NCI Cancer Post-GWAS Initiative (U19-CA148112). In addition, we acknowledge the following: AUS: U.S. Army Medical Research and Materiel Command (DAMD17-01-1- 0729), National Health & Medical Research Council of Australia, Cancer Councils of New South Wales, Victoria, Queensland, South Australia and Tasmania, Cancer Foundation of Western Australia; National Health and Medical Research Council of Australia (199600 and 400281). The Australian Ovarian Cancer Study Management Group (D. Bowtell, G. Chenevix-Trench, A. deFazio, D. Gertig, A. Green, P. Webb) and ACS Investigators (A. Green, P. Parsons, N. Hayward, P. Webb, D. Whiteman) thank all the clinical and scientific collaborators (see http://www.aocstudy.org/) and the women for their contribution.

GCT & PW are supported by Fellowships from NHMRC; AP is funded by a Medical Research Council studentship; DOV: U.S. National Institutes of Health R01-CA112523 and R01-CA87538; HAW: U.S. National Institutes of Health (R01- CA58598, N01-CN-55424 and N01-PC-67001); HOP: DOD DAMD17-02-1-0669 and NCI K07-CA080668, R01-CA95023, P50-CA159981; NIH/National Center for Research Resources/General Clinical Research Center grant M01-RR000056; R01-CA126841; JPN: Grant-in-Aid for the Third Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare; LAX: American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the L & S Milken Foundation; MAL: Funding for this study was provided by research grant R01- CA61107 from the National Cancer Institute, Bethesda, MD; research grant 94 222 52 from the Danish Cancer Society, Copenhagen, Denmark; and the Mermaid I project.; MAC and MAY: National Institutes of Health (R01-CA122443, P30-CA15083, P50-CA136393), Mayo Foundation; Minnesota Ovarian Cancer Alliance; Fred C. and Katherine B. Andersen Foundation; MAS: Malaysian Ministry of Higher Education (UM.C/HIR/MOHE/06) and Cancer Research Initiatives Foundation; NCO: National Institutes of Health (R01-CA76016) and the Department of Defense (DAMD17-02-1-0666); NEC: National Institutes of Health R01-CA54419 and P50-CA105009 and Department of Defense W81XWH- 10-1-02802; NHS: National Institute of Health (P01-CA87696 and R01-CA49449); ORE: OHSU Foundation; POC: Pomeranian Medical University; NJO: National Cancer Institute (NIH-K07 CA095666, R01-CA83918, NIH-K22-CA138563, and P30-CA072720), the Cancer Institute of New Jersey, and NCI CCSG award (P30- CA008748). POL: Intramural Research Program of the NCI; RMH: Cancer Research UK (no grant number is available); OVA: (MOP-86727, MSH-87734; SEA: Cancer Research UK (C490/A10119 C490/A10124); UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge, SEARCH team, Craig Luccarini, Caroline Baynes, Don Conroy; SIS: National Institute of Environmental Health Sciences, (Z01ES044005); SWH: National Institute for Health (R37-CA070867); UKO: The UKOPS

study was funded by The Eve Appeal (The Oak Foundation) and supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. We particularly thank I. Jacobs, M. Widschwendter, E. Wozniak, A. Ryan, J. Ford and N. Balogun for their contribution to the study; UCI: NIH R01-CA058860, NIH R01-CA092044, US Public Health Service PSA-042205, and the Lon V Smith Foundation grant LVS-39420; USC: P01-CA17054, P30-CA14089, R01-CA61132, N01-PC67010, R03-CA113148, R03-CA115195, N01-CN025403, and California Cancer Research Program (00-01389V-20170, 2II0200).

CONFLICTS OF INTEREST

The authors do not have any conflicts of interest to report.

REFERENCES

1 Stratton, J.F., Pharoah, P., Smith, S.K., Easton, D. and Ponder, B.A. (1998) A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br. J. Obstet. Gynaecol.*, **105**, 493-499.

2 Antoniou, A.C. and Easton, D.F. (2006) Risk prediction models for familial breast cancer. *Future Oncol.*, 2, 257-274.

3 Pharoah, P.D., Antoniou, A.C., Easton, D.F. and Ponder, B.A. (2008) Polygenes, risk prediction, and targeted prevention of breast cancer. *N. Engl. J. Med.*, **358**, 2796-2803.

4 Song, H., Ramus, S.J., Tyrer, J., Bolton, K.L., Gentry-Maharaj, A., Wozniak, E., Anton-Culver, H., Chang-Claude, J., Cramer, D.W., DiCioccio, R. *et al.* (2009) A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat. Genet.*, **41**, 996-1000.

5 Goode, E.L., Chenevix-Trench, G., Song, H., Ramus, S.J., Notaridou, M., Lawrenson, K., Widschwendter, M., Vierkant, R.A., Larson, M.C., Kjaer, S.K. *et al.* (2010) A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.*, in press.

6 Bolton, K.L., Tyrer, J., Song, H., Ramus, S.J., Notaridou, M., Jones, C., Sher, T., Gentry-Maharaj, A., Wozniak, E., Tsai, Y.Y. *et al.* (2010) Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat. Genet.*, in press.

7 Pharoah, P.D.P., Tsai, Y.-Y., Ramus, S.J., Phelan, C.M., Goode, E.L., Lawrenson, K., Buckley, M., Fridley, B.L., Tyrer, J.P., Shen, H. *et al.* (2013) GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.*, **45**, 362-370.

8 Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E. *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.*, **45**, 371-384, 384e371-372.

9 Permuth-Wey, J., Lawrenson, K., Shen, H.C., Velkova, A., Tyrer, J.P., Chen, Z., Lin, H.Y., Chen, Y. Ann, Tsai, Y.Y., Qu, X. *et al.* (2013) Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. *Nat. Commun.*, **4**, 1627.

10 Shen, H., Fridley, B.L., Song, H., Lawrenson, K., Cunningham, J.M., Ramus, S.J., Cicek, M.S., Tyrer, J., Stram, D., Larson, M.C. *et al.* (2013) Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. *Nat. Commun.*, **4**, 1628.

11 Chen, K., Ma, H., Li, L., Zang, R., Wang, C., Song, F., Shi, T., Yu, D., Yang, M., Xue, W. *et al.* (2014) Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. *Nat. Commun.*, **5**, 4682.

12 Kuchenbaecker, K.B., Ramus, S.J., Tyrer, J., Lee, A., Shen, H.C., Beesley, J., Lawrenson, K., McGuffog, L., Healey, S., Lee, J.M. *et al.* (2015) Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat. Genet.*, in press.

13 Kelemen, L.E., Lawrenson, K., Tyrer, J., Li, Q., Lee, J.M., Seo, J.H., Phelan, C.M., Beesley, J., Chen, X., Sprindler, T.J., Aben, K.K., Anton-Culver, H., Antonenkova, N., Australian Cancer Study, Australian Ovarian Cancer Study Group, Ovarian Cancer Associaton Constorium. (2015) Genome-wide significant risk associations for mucinous ovarian carcinoma. *Nat. Genet.*, **47**, 888-897.

14 Barrett, J.C. and Cardon, L.R. (2006) Evaluating coverage of genome-wide association studies. *Nat. Genet.*, **38**, 659-662.

15 Kryukov, G.V., Pennacchio, L.A. and Sunyaev, S.R. (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.*, **80**, 727-739.

16 Martin, A.R., Tse, G., Bustamante, C.D. and Kenny, E.E. (2014) Imputation-based assessment of next generation rare exome variant arrays. *Pac. Symp. Biocomput.*, in press., 241-252.

17 Abecasis G., Altshuler D., Boehnke M., Daly M., McCarthy M., Nickerson D., and Rich S.. Exome Chip Design. http://genome.sph.umich.edu/wiki/Exome_Chip_Design (Accessed April 16, 2014). in press.

18 Pirie, A., Wood, A., Lush, M., Tyrer, J. and Pharoah, P.D. (2015) The effect of rare variants on inflation of the test statistics in case-control analyses. *BMC Bioinformatics*, **16**, 53.

19 Pharoah, P.D., Tsai, Y.Y., Ramus, S.J., Phelan, C.M., Goode, E.L., Lawrenson, K., Buckley, M., Fridley, B.L., Tyrer, J.P., Shen, H. *et al.* (2013) GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.*, **45**, 362-370.

20 T.R., Kim, Y., Michael, A.F., Butkowski, R.J. and Charonis, A.S. (1998) Tubulointerstitial nephritis antigen (TIN-ag) is expressed in distinct segments of the developing human nephron. *Connect Tissue Res.*, **37**, 53-60.

21 Jia, D., Dong, R., Jing, Y., Xu, D., Wang, Q., Chen, L., Li, Q., Huang, Y., Zhang, Y., Zhang, Z. *et al.* (2014) Exome sequencing of hepatoblastoma reveals novel mutations and cancer genes in the Wnt pathway and ubiquitin ligase complex. *Hepatology*, **60**, 1686-1696.

22 Yang, L., Cui, Y., Shen, J., Lin, F., Wang, X., Long, M., Wei, J. and Zhang, H. (2015) Antitumor activity of SA12, a novel peptide, on SKBr-3 breast cancer cells via the mitochondrial apoptosis pathway. *Drug Des. Devel. Ther.*, **9**, 1319-1330.

23 Song, H., Dicks, E., Ramus, S.J., Tyrer, J.P., Intermaggio, M.P., Hayward, J., Edlund, C.K., Conti, D., Harrington, P., Fraser, L. *et al.* (2015) Contribution of Germline Mutations in the RAD51B, RAD51C, and RAD51D Genes to Ovarian Cancer in the Population. *J. Clin. Oncol.*, **33**, 2901-2907.

24 Ramus, S.J., Song, H., Dicks, E., Tyrer, J.P., Rosenthal, A.N., Intermaggio, M.P., Fraser, L., Gentry-Maharaj, A., Hayward, J., Philpott, S. *et al.* (2015) Germline Mutations in the BRIP1, BARD1, PALB2, and NBN Genes in Women With Ovarian Cancer. *J. Natl. Cancer Inst.*, 107.

25 Kanchi, K.L., Johnson, K.J., Lu, C., McLellan, M.D., Leiserson, M.D., Wendl, M.C., Zhang, Q., Koboldt, D.C., Xie, M., Kandoth, C. *et al.* (2014) Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.*, **5**, 3156.

26 http://www.ncbi.nlm.nih.gov/clinvar/, Accessed July 24, 2015. in press.

27 Ray, A.M., Zuhlke, K.A., Levin, A.M., Douglas, J.A., Cooney, K.A. and Petros, J.A. (2009) Sequence variation in the mitochondrial gene cytochrome c oxidase subunit I and prostate cancer in African American men. *Prostate*, **69**, 956-960.

28 Tyrer, J., Pharoah, P.D. and Easton, D.F. (2006) The admixture maximum likelihood test: a novel experiment-wise test of association between disease and multiple SNPs. *Genet. Epidemiol.*, **30**, 636-643.

29 Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, **89**, 82-93.

30 Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U S A*, **100**, 9440-9445.

31 Page, C.M., Baranzini, S.E., Mevik, B.H., Bos, S.D., Harbo, H.F. and Andreassen, B.K. (2015) Assessing the Power of Exome Chips. *PLoS One*, **10**, e0139642.

32 Clark, A.J., McLoughlin, L. and Grossman, A. (1993) Familial glucocorticoid deficiency associated with point mutation in the adrenocorticotropin receptor. *Lancet*, **341**, 461-462.

33 Hodebeck, M., Scherer, C., Wagner, A.H., Hecker, M. and Korff, T. (2014) TonEBP/NFAT5 regulates ACTBL2 expression in biomechanically activated vascular smooth muscle cells. *Front Physiol.*, **5**, 467.

34 Kang, U.B., Ahn, Y., Lee, J.W., Kim, Y.H., Kim, J., Yu, M.H., Noh, D.Y. and Lee, C. (2010) Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. *BMC Cancer*, **10**, 114.

35 So, A.K., Kaur, J., Kak, I., Assi, J., MacMillan, C., Ralhan, R. and Walfish, P.G. (2012) Biotinidase is a novel marker for papillary thyroid cancer aggressiveness. *PLoS One*, **7**, e40956.

36 Huang, L., Zheng, M., Zhou, Q.M., Zhang, M.Y., Jia, W.H., Yun, J.P. and Wang, H.Y. (2011) Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma. *Cancer*, **117**, 3363-3373.

37 Hernandez-Vazquez, A., Wolf, B., Pindolia, K., Ortega-Cuellar, D., Hernandez-Gonzalez, R., Heredia-Antunez, A., Ibarra-Gonzalez, I. and Velazquez-Arellano, A. (2013) Biotinidase knockout mice show cellular energy deficit and altered carbon metabolism gene expression similar to that of nutritional biotin deprivation: clues for the pathogenesis in the human inherited disorder. *Mol. Genet. Metab.*, **110**, 248-254.

38 Aspuria, P.J., Lunt, S.Y., Varemo, L., Vergnes, L., Gozo, M., Beach, J.A., Salumbides, B., Reue, K., Wiedemeyer, W.R., Nielsen, J. *et al.* (2014) Succinate dehydrogenase inhibition leads to epithelial-mesenchymal transition and reprogrammed carbon metabolism. *Cancer Metab.*, **2**, 21.

39 Kelemen, L.E., Goodman, M.T., McGuire, V., Rossing, M.A., Webb, P.M., Kobel, M., Anton- Culver, H., Beesley, J., Berchuck, A., Brar, S. *et al.* (2010) Genetic variation in TYMS in the one-carbon transfer pathway is associated with ovarian carcinoma types in the Ovarian Cancer Association Consortium. *Cancer Epidemiol. Biomarkers Prev.*, **19**, 1822-1830.
40 Walker, G., MacLeod, K., Williams, A.R., Cameron, D.A., Smyth, J.F. and Langdon, S.P. (2007) Estrogen-regulated gene expression predicts response to endocrine therapy in patients with ovarian cancer. *Gynecol. Oncol.*, **106**, 461-468.
41 Sheng, S., Barnett, D.H. and Katzenellenbogen, B.S. (2008) Differential estradiol and selective estrogen receptor modulator (SERM) regulation of Keratin 13 gene expression and its underlying mechanism in breast cancer cells. *Mol. Cell Endocrinol.*, 296, 1-9.
42 Hafiz, S., Dennis, J.C., Schwartz, D., Judd, R., Tao, Y.X., Khazal, K., Akingbemi, B., Mo, X.L., Abdel-Mageed, A.B., Morrison, E. *et al.* (2012) Expression of melanocortin receptors in human prostate cancer cell lines: MC2R activation by ACTH increases prostate cancer cell proliferation. *Int. J. Oncol.*, **41**, 1373-1380.
43 Hofland, J., Delhanty, P.J., Steenbergen, J., Hofland, L.J., van Koetsveld, P.M., van Nederveen, F.H., de Herder, W.W., Feelders, R.A. and de Jong, F.H. (2012) Melanocortin 2 receptor-associated protein (MRAP) and MRAP2 in human adrenocortical tissues: regulation of expression and association with ACTH responsiveness. *J. Clin. Endocrinol. Metab.*, **97**, E747-754.
44 Al Ojaimi, E.H. (2014) Cushing's syndrome due to an ACTH-producing primary ovarian carcinoma. *Hormones (Athens)*, **13**, 140-145.
45 Huang, B., Wu, X., Zhou, Q., Hu, Y., Zhao, H., Zhu, H., Zhang, Q. and Zheng, F. (2014) Cushing's syndrome secondary to ectopic ACTH secretion from carcinoid tumor within an ovarian mature teratoma: a case report and review of the literature. *Gynecol. Endocrinol.*, **30**, 192-196.
46 Yilmaz-Agladioglu, S., Savas-Erdeve, S., Boduroglu, E., Onder, A., Karaman, I., Cetinkaya, S. and Aycan, Z. (2013) A girl with steroid cell ovarian tumor misdiagnosed as non-classical congenital adrenal hyperplasia. *Turk. J. Pediatr.*, **55**, 443-446.
47 Maeda, O., Shibata, K., Hosono, S., Fujiwara, S., Kajiyama, H., Ino, K., Nawa, A., Tamakoshi, K. and Kikkawa, F. (2012) Spectrin alphaII and betaII tetramers contribute to platinum anticancer drug resistance in ovarian serous adenocarcinoma. *Int. J. Cancer*, **130**, 113-121.
48 Zhi, X., Lin, L., Yang, S., Bhuvaneshwar, K., Wang, H., Gusev, Y., Lee, M.H., Kallakury, B., Shivapurkar, N., Cahn, K. *et al.* (2015) betaII-Spectrin (SPTBN1) suppresses progression of hepatocellular carcinoma and Wnt signaling by regulation of Wnt inhibitor kallistatin. *Hepatology*, **61**, 598-612.
49 Jiang, X., Gillen, S., Esposito, I., Giese, N.A., Michalski, C.W., Friess, H. and Kleeff, J. (2010) Reduced expression of the membrane skeleton protein beta1-spectrin (SPTBN1) is associated with worsened prognosis in pancreatic cancer. *Histol. Histopathol.*, **25**, 1497-1506.
50 Lu, Y., Gu, X., Chen, L., Yao, Z., Song, J., Niu, X., Xiang, R., Cheng, T., Qin, Z., Deng, W. *et al.* (2014) Interferon-gamma produced by tumor-infiltrating NK cells and CD4+ T cells downregulates TNFSF15 expression in vascular endothelial cells. *Angiogenesis*, **17**, 529-540.
51 Holmen, O.L., Zhang, H., Zhou, W., Schmidt, E., Hovelson, D.H., Langhammer, A., Lochen, M.L., Ganesh, S.K., Mathiesen, E.B., Vatten, L. *et al.* (2014) No large-effect low-frequency coding variation found for myocardial infarction. *Hum. Mol. Genet.*, in press.
52 Chung, S.J., Kim, M.J., Kim, J., Kim, Y.J., You, S., Koh, J., Kim, S.Y. and Lee, J.H. (2014) Exome array study did not identify novel variants in Alzheimer's disease. *Neurobiol. Aging*, 35, 1958 e1913-1954.
53 Huyghe, J.R., Jackson, A.U., Fogarty, M.P., Buchkovich, M.L., Stancakova, A., Stringham, H.M., Sim, X., Yang, L., Fuchsberger, C., Cederberg, H. *et al.* (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.*, **45**, 197-201.
54 Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76-82.
55 Fasching, P.A., Gayther, S., Pearce, L., Schildkraut, J.M., Goode, E., Thiel, F., Chenevix-Trench, G., Chang-Claude, J., Wang-Gohrke, S., Ramus, S. *et al.* (2009) Role of genetic polymorphisms and ovarian cancer susceptibility. *Mol. Oncol.*, **3**, 171-181.

56 Grove, M.L., Yu, B., Cochran, B.J., Haritunians, T., Bis, J.C., Taylor, K.D., Hansen, M., Borecki, I.B., Cupples, L.A., Fornage, M. *et al.* (2013) Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One*, 8, e68095.
57 Sankararaman, S., Sridhar, S., Kimmel, G. and Halperin, E. (2008) Estimating local ancestry in admixed populations. *Am. J. Hum. Genet.*, **82**, 290-303.
58 Oliphant, A., Barker, D.L., Stuelpnagel, J.R. and Chee, M.S. (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques*, Suppl, 56-58, 60-51.
59 Goldstein, J.I., Crenshaw, A., Carey, J., Grant, G.B., Maguire, J., Fromer, M., O'Dushlaine, C., Moran, J.L., Chambert, K., Stevens, C. *et al.* (2012) zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics*, **28**, 2543-2545.
60 Robertson, A. and Hill, W.G. (1984) Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics*, **107**, 703-718.
61 Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N. *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nat. Genet.*, **36**, 388-393.
62 Lehne, B., Lewis, C.M. and Schlitt, T. (2011) Exome localization of complex disease association signals. *BMC Genomics*, **12**, 92.
63 Tyrer, J.P., Guo, Q., Easton, D.F. and Pharoah, P.D. (2013) The admixture maximum likelihood test to test for association between rare variants and disease phenotypes. *BMC Bioinformatics*, **14**, 177.
64 Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., Christiani, D.C., Wurfel, M.M. and Lin, X. (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.*, **91**, 224-237.
65 Cancer Genome Atlas Research Network,.(2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, 474, 609-615.

LEGENDS TO FIGURES

Figure 1. Manhattan plot of association for 98,299 SNPs from a pooled analysis of Affymetrix and Illumina exome genotyping arrays. Plots show the strength of association versus chromosomal position for (A) all invasive EOC risk and (B) serous invasive EOC risk. The red line represents exome-wide significance ($5.0x10^{-7}$). Exome-wide significant variants are annotated for the gene in which they are located. Known variants previously reported to have the strongest association signal are indicated by a black diamond.

Figure 2. Manhattan plot of association for sub-exome-wide ($P \geq 5.0x10^{-7}$) variants from a pooled analysis of Affymetrix and Illumina exome genotyping arrays. SNPs with $P < 5.0x10^{-7}$ were filtered out and the strength of genetic association versus chromosomal position was plotted for the remaining 98,287 SNPs for the risk of (A) all invasive EOCs, (B) serous invasive EOCs, and (C) endometrioid invasive EOCs. Known variants previously reported to have the strongest association signal are indicated by a black diamond. Sub-exome-wide significant SNPs ($P < 5.0x10^{-5}$) are annotated for the gene in which they are located.

Figure 3. Regional association plots for rare variants associated with EOC susceptibility.
A) *BTD* rs200337373 (3p25.1), B) *ACTBL2* rs73757391 (5q11.2), C) *KRT13* rs150321809, (17q21.2), and D) *MC2R* rs104894658 (18p11.21)). Linkage disequilibrium (LD, r^2) between the strongest signal (noted by a purple diamond) and other variants is indicated by the color scheme.

Figure 4. Gene-level association of rare variants (MAF<1%) using the Rare Admixture Maximum Likelihood (RAML) association test. Results of association with all invasive EOC risk are shown for 15,118 genes, adjusting for study and first 5 PCs. Genes with $P < 5.0x10^{-4}$ are annotated.

Supplemental Figure 1. Sample and variant quality control and data pooling flowchart.

Supplemental Figure 2. Quantile-Quantile plot of distribution of test statistics for 98,299 SNPs in primary pooled analysis. Plot shows the observed test statistics from the likelihood ratio test versus the

expected test statistic. The ratio of the empirically observed median test statistic to the expected median (λ) is 1.13.

Supplemental Figure 3. Regional association plot for variants directly genotyped or imputed at the 3q25 susceptibility locus. Directly genotyped variant in *LEKR1*, rs62273959, is strongly correlated ($0.93>r^2>0.96$) with previously reported intronic variant near *TIPARP* (rs7651446) and newly imputed variant, rs78561123,which that maps within 0.5kb of the 3'UTR for *LINC00886*.

Supplemental Figure 4. Manhattan plot of association for 80,178 SNPs from Set 1 only analysis (Affymetrix). Plots show association between genotype and risk of (A) all invasive EOC, (B) serous EOC, and (C) endometrioid EOC. The red line represents exome-wide significance (5.0x10-7) and exome-wide or sub-exome-wide significant SNPs ($P < 5.0x10^{-5}$) are annotated for the gene in which they are located.

Supplemental Figure 5. Gene-level association of rare variants (MAF<1%) using RAML association test for Set 1 only analysis (Affymetrix). Results of association with all invasive EOC risk are shown for 15,042 genes, adjusting for the first 3 PCs. Genes with $P < 5.0x10^{-4}$ are annotated.

Table 1. Variants at Novel Loci Associated with Epithelial Ovarian Cancer Susceptibility with $5.0x10^{-5}>P>=5.0\ x10^{-7}$

							All Invasive (7308 cases, 10773 controls)		Serous (5955 cases, 10773 controls)		Endometrioid (883 cases, 10773 controls)	
Region[a]	rsID (major>minor allele)	Position (hg19)	Nearest Gene(s)	Function (Polyphen score/ prediction)	Case MAF	Control MAF	OR (95% CI)[b]	P value	OR (95% CI)[b]	P value	OR (95% CI)[b]	P value
1p36.33	rs138031468 (G>T)	977028	AGRN	Missense: A375S	0.005	0.007	0.79 (0.60-1.03)	8.34E-02	0.91 (0.7-1.2)	5.20E-01	0.08 (0.01-0.56)	3.80E-05
2p22.1	rs61757604 (C>T)	39095403	DHX57	Missense: G49S (0.318/B)	0.02	0.02	0.82 (0.68-0.99)	3.55E-02	0.92 (0.75-1.11)	3.69E-01	0.26 (0.12-0.56)	2.21E-05
3p25.1	rs200337373 (G>A)	15686027	BTD	Missense: D222N (0.999/D)	0.001	0	NE[c]	8.61E-06	NE[c]	1.75E-05	NE[c]	4.74E-02
3p14.2	rs4679621 (C>T)	59324733	C3orf67 (289kb) FHIT (410kb)	Intergenic	0.47	0.44	1.1 (1.05-1.14)	4.19E-05	1.1 (1.05-1.15)	3.95E-05	1.03 (0.93-1.14)	5.82E-01
5q11.2	rs381852 (G>A)	54459961	GPX8 CDC20B	Missense: K182R (0/B), Intron	0.21	0.18	1.13 (1.07-1.2)	1.63E-05	1.14 (1.08-1.21)	9.07E-06	1.05 (0.92-1.2)	4.72E-01
5q11.2	rs73757391 (C>T)	56778213	ACTBL2	Missense: E108K (1.0/D)	0.001	0	NE[c]	6.37E-06	NE[c]	5.15E-06	NE[c]	4.24E-02
6p22.1	rs114979098 (C>T)	29785235	HLA-G (10kb)	Intergenic	0.42	0.41	1.11 (1.05-1.18)	4.82E-04	1.08 (1.02-1.15)	1.15E-02	1.37 (1.19-1.56)	5.15E-06
6p21.33	rs149771958 (C>T)	31079994	C6orf15	Missense: G48R (0.895/PD)	0.07	0.08	0.84 (0.77-0.91)	4.24E-05	0.84 (0.77-0.92)	1.25E-04	0.86 (0.71-1.03)	1.01E-01
	rs116682468 (C>T)	31112484	CCHCR1	Missense: R627Q (1.0/D)	0.07	0.07	0.83 (0.76-0.91)	3.90E-05	0.84 (0.77-0.92)	2.86E-04	0.82 (0.68-1)	4.55E-02
	rs116151586 (T>C)	31118019	CCHCR1	Intron	0.32	0.32	0.88 (0.83-0.94)	4.60E-05	0.88 (0.83-0.94)	9.03E-05	0.84 (0.73-0.97)	1.39E-02
	rs114470046 (C>A)	31125257	CCHCR1	Nonsense: E41Stop	0.07	0.07	0.83 (0.76-0.91)	3.61E-05	0.84 (0.77-0.92)	2.75E-04	0.82 (0.67-1)	4.31E-02
	rs115538919 (C>T)	31129707	TCF19	Missense: P241L (0.906/PD)	0.07	0.07	0.83 (0.76-0.91)	3.76E-05	0.84 (0.77-0.93)	3.09E-04	0.82 (0.68-1)	4.49E-02
6p21.33	rs113935384 (G>A)	31231989	HCG27 (60kb) HLA-C (4.5kb)	Intergenic	0.36	0.36	0.89 (0.84-0.94)	3.66E-05	0.89 (0.84-0.94)	1.09E-04	0.9 (0.79-1.02)	9.54E-02
6p12.1	rs2297980 (A>G)	54173413	TINAG	Missense: Q22R (0/B)	0.09	0.11	0.84 (0.78-0.9)	1.90E-06	0.83 (0.77-0.9)	2.25E-06	0.91 (0.76-1.07)	2.47E-01
6q25.2	rs199761238 (T>C)	152652052	SYNE1	Missense: N4590D (0.818/PD)	0	0.001	0.15 (0.03-0.63)	1.06E-03	NE[c]	2.78E-05	1.33 (0.29-6.08)	7.22E-01
11q13.1	rs145514333 (C>T)	64527189	PYGM	Missense: R61H (1.0/D)	0.004	0.0001	3.18 (1.69-6.01)	1.49E-04	3.59 (1.91-6.77)	2.67E-05	0.84 (0.19-3.77)	8.18E-01
15q12	rs147432497 (G>A)	25940059	ATP10A	Missense: R999C (1.0/D)	0	0.001	0.16 (0.04-0.67)	1.61E-03	NE[c]	4.13E-05	NE[c]	1.31E-01
16q22.3	rs147445846 (G>C)	72992910	ZFHX3	Missense: L379V (0.979/D)	0.001	0.002	0.37 (0.22-0.62)	4.86E-05	0.34 (0.19-0.61)	7.01E-05	0.42 (0.13-1.34)	9.42E-02
17q21.2	rs150321809 (C>T)	39657599	KRT13	Missense: R429H	0.003	0.001	2.24 (1.52-3.31)	3.20E-05	2.45 (1.65-3.65)	6.49E-06	1.03 (0.4-2.67)	9.48E-01

				(0.266/B)								
18p11.21	rs104894658 (C>A)	13885297	MC2R	Missense: S74I (1.0/D)	0.002	0.0003	9.66 (2.73-34.24)	3.66E-05	10.15 (2.86-35.99)	2.76E-05	NE[c]	5.78E-02
22q11.2	rs141200301 (C>T)	24123521	MMP11	Missense: R334C (1.0/D)	0.001	0	NE[c]	7.49E-05	NE[c]	3.91E-05	NA	NA

Table 2: Genes most strongly associated with epithelial ovarian cancer risk, with P<5.0x10^{-4} by RAML

Gene Information			Rare Admixture Maximum Likelihood (RAML) test[1]						Sequence Kernel Association Test (SKAT)[1]					
			All Invasive		Serous		Endometrioid		All Invasive		Serous		Endometrioid	
Gene	Region	N SNPs (total/rare)	P-value[2]	FDR	P-value[2]	FDR	P-value[2]	FDR	P-value[2]	FDR	P-value[2]	FDR	P-value[2]	FDR
ACTBL2	5q11.2	6/6	3.23E-05	0.38	**2.00E-05**	**0.15**	2.99E-01	1	9.23E-04	0.56	1.50E-03	0.46	2.60E-01	0.98
KRT13	17q21.2	7/7	1.67E-04	0.38	**2.94E-05**	**0.15**	2.65E-01	1	**2.04E-05**	**0.15**	**5.24E-06**	**0.08**	3.62E-01	0.98
CLCNKA	1p36.13	3/3	2.00E-04	0.38	8.33E-05	0.25	5.60E-01	1	2.98E-04	0.48	**6.08E-05**	**0.14**	5.61E-01	0.98
MYO19	17q12	14/11	2.00E-04	0.38	7.50E-04	0.37	1.84E-02	1	5.41E-03	0.64	9.30E-03	0.66	1.76E-02	0.74
TNFSF15	9q32	8/3	2.20E-04	0.38	5.56E-05	0.21	2.56E-01	1	2.79E-04	0.48	**4.39E-05**	**0.14**	2.62E-01	0.98
TRIB1	8q24.13	4/2	2.80E-04	0.38	3.67E-04	0.29	7.76E-01	1	4.90E-04	0.51	3.09E-04	0.26	7.25E-01	0.99
MC2R	18p11.21	6/6	3.25E-04	0.38	2.00E-04	0.29	9.62E-02	1	1.32E-01	0.91	2.01E-01	0.93	8.26E-02	0.98
LIG3	17q12	7/6	3.33E-04	0.38	1.70E-03	0.46	3.59E-02	1	4.91E-03	0.64	9.80E-03	0.66	4.62E-03	0.55
CAMSAP3	19p13.2	4/3	3.67E-04	0.38	2.40E-04	0.29	3.54E-01	1	1.10E-04	0.37	**6.76E-05**	**0.14**	4.20E-01	0.98
GSDMB	17q21.1	15/7	3.67E-04	0.38	4.33E-04	0.31	5.41E-01	1	2.78E-01	0.95	2.47E-01	0.93	2.49E-01	0.98
KIAA1586	6p12.1	3/2	4.00E-04	0.38	3.33E-04	0.29	1.80E-01	1	4.70E-04	0.51	2.00E-04	0.26	1.53E-01	0.98
SPTBN1	2p16.2	9/8	4.00E-04	0.38	2.10E-03	0.48	2.06E-01	1	**1.16E-05**	**0.15**	2.20E-04	0.26	3.88E-02	0.97
STPG1	1p36.11	9/3	4.50E-04	0.40	7.50E-04	0.37	2.21E-01	1	4.50E-01	0.95	1.71E-01	0.92	1.84E-01	0.98
GPATCH2	1q41	8/7	5.00E-04	0.42	3.67E-04	0.29	7.06E-01	1	3.26E-02	0.81	5.37E-02	0.84	1.00E+00	1.00
CCDC136	7q33	12/11	5.33E-04	0.42	1.43E-04	0.26	7.47E-01	1	1.07E-02	0.75	8.68E-03	0.66	7.53E-01	0.99
WDR59	16q23.1	10/8	1.00E-03	0.50	3.67E-04	0.29	2.58E-01	1	2.81E-03	0.61	2.67E-03	0.55	4.51E-01	0.98
NEXN	1p31.1	3/3	1.00E-03	0.50	3.67E-04	0.29	1.91E-01	1	5.46E-03	0.64	3.64E-03	0.61	4.53E-01	0.98
MMP11	22q11.23	5/4	1.20E-03	0.50	4.33E-04	0.31	2.98E-01	1	2.69E-01	0.95	4.25E-02	0.83	1.87E-01	0.98
BCL9L	11q23.3	2/2	1.30E-03	0.50	1.57E-04	0.26	4.52E-01	1	1.31E-03	0.57	1.58E-04	0.26	4.99E-01	0.98
PYGM	11q12-q13.2	11/11	3.10E-03	0.66	3.33E-04	0.29	3.52E-01	1	1.56E-01	0.92	5.24E-02	0.83	3.13E-01	0.98
ATMIN	16q23.2	7/6	5.06E-02	0.94	3.27E-01	1	3.67E-04	1	1.23E-01	0.91	5.30E-01	0.94	9.09E-04	0.30
SOS2	14q21	9/9	1.30E-01	0.99	3.71E-01	1	6.88E-05	1	8.85E-02	0.89	1.83E-01	0.92	6.34E-02	0.98
OR7G1	19p13.2	9/3	1.97E-01	1.00	3.85E-01	1	2.40E-04	1	9.47E-02	0.90	3.00E-01	0.94	4.23E-04	0.23
PRR5	22q13	5/4	3.71E-01	1.00	6.75E-01	1	3.67E-04	1	4.08E-01	0.95	8.36E-01	0.94	7.86E-02	0.98

[1]Both the RAML and SKAT methods were limited to rare variants (MAF<1%). No weighting by minor allele frequency was used in either method.

[2]Analyses are adjusted for site and the first five principal components representing European ancestry.