Inherited variants affecting RNA editing may contribute to ovarian cancer susceptibility: results from a large-scale collaboration

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.18632/oncotarget.10546</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:32072222">http://nrs.harvard.edu/urn-3:HUL.InstRepos:32072222</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Inherited variants affecting RNA editing may contribute to ovarian cancer susceptibility: results from a large-scale collaboration


¹ Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
² Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA
³ Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA
⁴ Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Queensland, Australia
⁵ Cancer Genetics Laboratory, Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia
⁶ David Geffen School of Medicine, Department of Medicine, Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA
⁷ University Hospital Erlangen, Department of Gynecology and Obstetrics, Friedrich-Alexander-University Erlangen-Nuremberg, Comprehensive Cancer Center, Erlangen, Germany
⁸ Vesalius Research Center, VIB, Leuven, Belgium
⁹ Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium
¹⁰ Division of Gynecologic Oncology, Department of Obstetrics and Gynecology and Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium
¹¹ Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
¹² Department of Epidemiology, University of Washington, Seattle, WA, USA
¹³ Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, NY, USA
¹⁴ German Cancer Research Center, Division of Cancer Epidemiology, Heidelberg, Germany
¹⁵ University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
RNA editing in mammals is a form of post-transcriptional modification in which adenosine is converted to inosine by the adenosine deaminases acting on RNA (ADAR) family of enzymes. Based on evidence of altered ADAR expression in epithelial ovarian cancers (EOC), we hypothesized that single nucleotide polymorphisms (SNPs) in ADAR genes modify EOC susceptibility, potentially by altering ovarian tissue gene expression. Using directly genotyped and imputed data from 10,891 invasive EOC cases and 21,693 controls, we evaluated the associations of 5,303 SNPs in ADAD1, ADAR, ADAR2, ADAR3, and SND1. Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI), with adjustment for European ancestry. We conducted gene-level analyses using the Admixture Maximum Likelihood (AML) test and the Sequence-Kernel Association test for common and rare variants (SKAT-CR). Association analysis revealed top risk-associated SNP rs77027562 (OR (95% CI)= 1.39 (1.17-1.64), P=1.0x10^-4) in ADAR3 and rs185455523 in SND1 (OR (95% CI)= 0.68 (0.56-0.83), P=2.0x10^-4). When restricting to serous...
INTRODUCTION

Over the past decade it has been recognized that the complexity of higher organisms is related to the information stored in non-protein-coding regions of the genome. Such complexity may be attributed to a range of processing events and post-transcriptional modifications that affect the fate of RNA, including alternative splicing, 5’ capping, 3’ polyadenylation, and RNA editing [1-3]. The most common type of RNA editing in eukaryotes is site-selective hydrolytic deamination of adenosine into inosine (A-to-I) within double-stranded RNAs, and recent bioinformatic analyses and high-throughput sequencing efforts have revealed that A-to-I editing is widespread and alters non-coding and protein-coding sequences throughout the genome [4].

A-to-I editing is mediated by a family of adenosine deaminases acting on RNA (ADARs), and this process modulates expression of genes and biological pathways via several mechanisms [4]. Indeed, altered expression and/or activity of ADAR enzymes has been linked to a variety of conditions, including cardiovascular and neurological diseases and cancers [4]. Epithelial ovarian cancer (EOC) is the fifth leading cause of cancer death among women in the United States [5], and ADAR expression levels have been reported to be significantly higher in serum and peritoneal fluid from patients with EOCs compared with benign ovarian tumors [6, 7], suggesting ADARs may be useful biomarkers for the diagnosis and management of EOC.

We hypothesized that germline single nucleotide polymorphisms (SNPs) involving ADAR-related RNA editing genes may contribute to EOC risk. The main purpose of this investigation was to determine whether SNPs in five ADAR genes (ADAD1, ADAR, ADAR2, ADAR3, and SND1) were associated with EOC susceptibility. We used data available from a large-scale genotyping collaboration involving 10,891 EOC cases and 21,693 controls from the International Ovarian Cancer Association Consortium (OCAC) [8]. We also sought to evaluate the overall contribution of each gene on EOC susceptibility and to determine whether candidate SNPs associated with altered expression of corresponding genes in EOC tumor tissue.

RESULTS

Study population

The study sample included 10,891 invasive EOC patients and 21,693 controls of European ancestry (Supplementary Table 1). Selected subject characteristics are shown in Table 1. The mean age at diagnosis for cases was 58.1 years, the mean age at interview for controls was 56.1 years. Cases were more likely than controls to be nulliparous and to have never used oral contraceptives. Most cases had serous histology (59.7%), distant stage (63.0%), and high-grade disease (58.9%).

Variant-level association analysis and overlap with regulatory domains

SNP-level association analysis revealed top-ranked SNPs (defined as the top 5% of SNPs having the most statistically significant P values) in ADAR, ADAR3, and SND1 in the all-histologies and serous-only analyses (Figure 1A and 1B). Table 2 summarizes association results for the most statistically significant SNPs overall or by serous histology (P < 4.0x10⁻⁸); associations were not significant after correction for multiple testing (FDR > 0.15). Most of the top-ranked variants were imputed, rare or low frequency (MAF < 0.05), and not part of a shared haplotype. rs77027562 (A>G; MAF = 0.009), the top risk-associated variant among all histologies (OR (95% CI) = 1.39 (1.17-1.64, P = 1.0 x10⁻⁴)), resides in an intron of ADAR3. ADAR3 SNP rs77027562 and its proxies (r²>0.80) reside in genomic regions that overlap with regulatory domains, particularly enhancers in blood and brain (Table 3). The next top-ranked variant, SND1 rs185455523 (T>A), was associated with a decreased EOC risk (OR (95% CI) = 0.68 (0.56-0.83), P = 1.5 x10⁻⁴), but this SNP and its proxies do not appear to overlap with regulatory domains. When analysis was restricted to the 6,500 patients with invasive serous adenocarcinomas, the magnitude of association was slightly attenuated for ADAR3 rs77027562 (OR = 1.33, P = 6.1x10⁻⁵) and slightly stronger for SND1 rs185455523 (OR = 0.60, P = 1x10⁻⁴). Exploratory analysis for the less common histologic subtypes (endometrioid (n = 1,439), mucinous...
Figure 1: Manhattan plot for candidate RNA editing SNPs among a) all invasive cases \((n = 10,891)\) versus controls \((n = 21,693)\) and b) serous cases \((n = 6,500)\) versus controls.
(n = 696), and clear cell (n = 660)) revealed several SNP-level associations unique to each sub-type (Figure 2A-2C). For example, rs145678553-C in ADAR3 is a rare variant (MAF = 0.0047) associated with an increased risk for mucinous EOC (OR (95%CI) = 3.46 (1.91-6.26), P = 3.99x 10^{-5}), and rs116983191-A in ADAR3 is a low-frequency variant (MAF = 0.044) associated with clear cell carcinoma (OR (95%CI) = 1.86 (1.42-2.43, P = 6.91x 10^{-6}). rs145678553-C was not represented in Haploreg.

rs116983191-A is located in promoter and enhancer regions, but not in tissues relevant to ovarian cancer.

Gene-level analyses

Gene-level analyses based on AML and SKAT-CR revealed that variation in ADAR was nominally associated (P < 0.05) with susceptibility to all invasive EOC, with P = 0.02 using both methods (Table 4). Histology-specific analyses revealed that ADAR variation was associated with endometrioid EOC susceptibility (P_{SKAT-CR} = 0.005/P_{AML} = 0.008). When using a Bonferroni threshold of 0.0025, only ADAR3 variation was significantly associated with mucinous carcinoma (P_{SKAT-CR} = 0.0016/P_{AML} = 0.031).

To examine associations between genotype and gene expression for the 5 candidate RNA editing genes, expression quantitative trait locus (eQTL) analysis was performed using matched genotype and tissue expression data from The Cancer Genome Atlas (TCGA) high-grade serous adenocarcinoma tumors (https://tcga-data.nci.nih.gov/tcga/). eQTL analysis revealed statistically significant associations (P < 0.05) with ADAR expression for several SNPs in ADAR, including rs1127313 (G/A), a SNP in the 3’UTR within a putative miRNA binding site that was associated with susceptibility in all histologies (OR = 1.05, P = 0.009). rs1127313 is also in high LD (r^2 = 0.86) with top ADAR risk SNP rs9426826 (see Table 2). ADAR tumor tissue expression was slightly higher among G allele carriers of rs1127313 compared to A allele carriers (P = 0.027; Figure 3). rs1127313 is also an eQTL for ADAR in whole blood (Supplementary Table 2), and lies in a genomic region with enhancer features and DNase I hypersensitivity site in several tissues, including ovary. Statistically-significant cis-eQTLs were not detected for SNPs in other candidate RNA editing genes.

DISCUSSION

An emerging body of data suggest that defects in RNA editing may contribute to a range of human diseases, including cancer [2-4, 9-11]. The current large-scale collaboration represents the first comprehensive association study of germline variants involving RNA editing genes and susceptibility to epithelial ovarian cancer. At the SNP-level, the strongest associations were observed for SNPs in RNA editing genes ADAR3 and SND1, but no associations reached genome-wide significance.
Figure 2: Manhattan plot for candidate RNA editing SNPs among a) endometrioid cases (n = 1,439) versus controls (n = 21,693), b) mucinous cases (n = 696) versus controls, and c) clear cell cases (n = 660) versus controls.
Table 2: Top-ranked RNA editing SNP-EOC risk associations among all histologies (N = 10,891) or serous histology (N = 6,500) versus controls (N = 21,693), sorted by gene and p-value

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Alleles</th>
<th>MAF</th>
<th>Imputation accuracy R2</th>
<th>All histologies OR (95% CI)</th>
<th>P</th>
<th>FDR</th>
<th>Serous OR (95% CI)</th>
<th>P</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAR</td>
<td>rs9426826</td>
<td>C&gt;G</td>
<td>0.481</td>
<td>0.86</td>
<td>1.05 (1.02-1.09)</td>
<td>0.0038</td>
<td>0.76</td>
<td>1.04 (1.08-1.10)</td>
<td>0.0759</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs3738030</td>
<td>A&gt;C</td>
<td>0.116</td>
<td>0.79</td>
<td>0.93 (0.89-0.98)</td>
<td>0.0080</td>
<td>0.83</td>
<td>0.91 (0.86-0.97)</td>
<td>0.0038</td>
<td>0.9996</td>
</tr>
<tr>
<td>ADAR3</td>
<td>rs77027562</td>
<td>A&gt;G</td>
<td>0.009</td>
<td>0.41</td>
<td>1.39 (1.17-1.64)</td>
<td>0.0001</td>
<td>0.34</td>
<td>1.33 (1.08-1.62)</td>
<td>0.0061</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs11250601</td>
<td>C&gt;T</td>
<td>0.070</td>
<td>0.62</td>
<td>0.89 (0.83-0.95)</td>
<td>0.0007</td>
<td>0.34</td>
<td>0.90 (0.83-0.97)</td>
<td>0.0071</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs142123280</td>
<td>A&gt;G</td>
<td>0.001</td>
<td>0.48</td>
<td>2.08 (1.36-3.17)</td>
<td>0.0007</td>
<td>0.34</td>
<td>2.01 (1.23-3.29)</td>
<td>0.0056</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs4880912</td>
<td>T&gt;C</td>
<td>0.200</td>
<td>0.82</td>
<td>1.07 (1.03-1.11)</td>
<td>0.0015</td>
<td>0.51</td>
<td>1.06 (1.01-1.11)</td>
<td>0.0267</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs1198359</td>
<td>C&gt;T</td>
<td>0.005</td>
<td>0.45</td>
<td>0.68 (0.54-0.87)</td>
<td>0.0018</td>
<td>0.53</td>
<td>0.63 (0.47-0.86)</td>
<td>0.0032</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs6560760</td>
<td>C&gt;T</td>
<td>0.025</td>
<td>0.59</td>
<td>1.17 (1.06-1.3)</td>
<td>0.0024</td>
<td>0.65</td>
<td>1.07 (0.95-1.22)</td>
<td>0.2688</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs2676202</td>
<td>C&gt;T</td>
<td>0.122</td>
<td>0.66</td>
<td>0.93 (0.88-0.98)</td>
<td>0.0038</td>
<td>0.76</td>
<td>0.9 (0.85-0.96)</td>
<td>0.0014</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs139646191</td>
<td>TAGAA&gt;T</td>
<td>0.062</td>
<td>0.66</td>
<td>1.11 (1.03-1.18)</td>
<td>0.0038</td>
<td>0.76</td>
<td>1.07 (0.98-1.16)</td>
<td>0.1242</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs139812582</td>
<td>G&gt;A</td>
<td>0.002</td>
<td>0.47</td>
<td>1.71 (1.15-2.54)</td>
<td>0.0078</td>
<td>0.83</td>
<td>2.03 (1.31-3.14)</td>
<td>0.0017</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>chr10:1419524</td>
<td>T&gt;GTT</td>
<td>0.009</td>
<td>0.60</td>
<td>0.79 (0.65-0.95)</td>
<td>0.0106</td>
<td>0.83</td>
<td>0.68 (0.54-0.86)</td>
<td>0.0014</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs185147330</td>
<td>C&gt;T</td>
<td>0.005</td>
<td>0.45</td>
<td>1.32 (1.05-1.67)</td>
<td>0.0176</td>
<td>0.86</td>
<td>1.54 (1.19-2.2)</td>
<td>0.0011</td>
<td>0.9996</td>
</tr>
<tr>
<td>SND1</td>
<td>rs185455523</td>
<td>T&gt;A</td>
<td>0.008</td>
<td>0.56</td>
<td>0.68 (0.60-0.83)</td>
<td>0.0002</td>
<td>0.34</td>
<td>0.6 (0.46-0.77)</td>
<td>0.0001</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>rs145106202</td>
<td>G&gt;C</td>
<td>0.009</td>
<td>0.87</td>
<td>0.73 (0.61-0.88)</td>
<td>0.0008</td>
<td>0.35</td>
<td>0.67 (0.53-0.84)</td>
<td>0.0006</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs181460088</td>
<td>C&gt;T</td>
<td>0.008</td>
<td>0.80</td>
<td>0.73 (0.60-0.89)</td>
<td>0.0021</td>
<td>0.58</td>
<td>0.67 (0.52-0.86)</td>
<td>0.0020</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs199750392</td>
<td>G&gt;GT</td>
<td>0.036</td>
<td>0.61</td>
<td>1.14 (1.05-1.25)</td>
<td>0.0028</td>
<td>0.71</td>
<td>1.11 (1.01-1.23)</td>
<td>0.0550</td>
<td>0.9996</td>
</tr>
<tr>
<td></td>
<td>rs79133832</td>
<td>C&gt;T</td>
<td>0.007</td>
<td>0.72</td>
<td>1.32 (1.09-1.61)</td>
<td>0.0056</td>
<td>0.83</td>
<td>1.42 (1.13-1.77)</td>
<td>0.0025</td>
<td>0.9996</td>
</tr>
</tbody>
</table>

Significant SNPs (P < 4.0x10^-6) are listed and SNPs in LD (r^2>0.60) with more significant SNP are not reported:

a) 7 SNPs in LD not reported
b) 12 SNPs in LD not reported
c) 1 SNP in LD not reported
d) 1 SNP in LD not reported

levels of statistical significance. Gene-level analyses highlighted ADAR and ADAR3 as potential contributors to EOC susceptibility within the set of ADAR-related genes. Finally, positive eQTLs were also observed between ADAR genotype and ADAR expression in EOC tumor tissue.

Focused evaluations of RNA editing SNP-disease associations are limited [12], especially with cancer as an outcome, so it is not possible to compare our SNP findings to those of other studies of cancer risk. We are, however, unaware of GWAS hits in or near these genes. Several recent studies [2, 3] have evaluated the genomic landscape and clinical relevance of RNA editing in numerous human tissue types. These analyses used RNA-sequencing data from both tumor and normal samples profiled as part of TCGA Project. Striking differences in RNA-editing patterns were observed in tumors relative to matched normal tissues for 12 cancer types [2]. Further analyses revealed that altered RNA editing patterns in tumors correlated with ADAR expression, and that non-random, clinically-relevant RNA editing events (frequently located in noncoding RNAs, nonsynonymous sites, intronic regions, and non-Alu elements) correlated with tumor classification and patient survival and with increased cell survival and altered drug sensitivity [2, 3]. Interestingly, gene amplification-associated overexpression of ADAR was recently shown to enhance lung tumorigenesis and contribute to poor outcomes by affecting downstream RNA editing patterns [10]. As mentioned previously, ADAR expression levels have been reported to be significantly higher in serum and/or peritoneal fluid from patients with EOCs compared with benign ovarian tumors [6, 7]. Although high-grade serous EOCs from TCGA were not profiled as part of the aforementioned genomic investigations [18, 19], Haploreg 4.1 effectively integrates GTEx eQTL results for normal ovary.

Taken together with several lines of investigation from ovarian [6, 7] and other cancers [2, 3, 10] the current study suggests that ADARs (and ADAR in particular) may be useful biomarkers for the diagnosis and management of EOC. Thus, with replication, ADAR genotype status and/or expression level may serve as a risk factor for EOC. Indeed, we find that our top risk SNP in ADAR, rs9426826, has several proxy variants (r^2>0.8, Supplementary Table 2) that are strongly associated with expression of this gene in blood (rs1127313: 7.23x10^{-14}) and to a lesser extent, expression in high-grade serous EOC tumors (rs1127313: P = 0.027). Based on growing data which demonstrate
the inhibition of tumor growth in the presence of ADAR inhibitors [13] and other therapeutic agents such as the IGFR-1R inhibitor BMS536924 and the MEK inhibitors CI1040 and trametinib [2], ADAR genotype and/or expression may help identify women whose tumors may respond to new combinations of therapies.

Strengths of the current study include the large sample size that primarily enabled detection of small effects for common variants, the relatively homogeneous population of EOC cases, and the multi-tiered genomic evaluation. However, this study was underpowered to detect the rare variants that were identified and is burdened by the low imputation quality. Additionally, the study is limited in that eQTL analysis did not permit adjustment for somatic copy number changes and DNA methylation status, factors that can influence transcript abundance and confound associations between germline polymorphisms and gene expression [14-16]. Moreover, it is possible that the top-ranked SNPs could potentially affect genes other than the RNA editing genes that drive candidate

Table 3: HaploReg results for top-ranked ADAR3 SNP rs77027562 and its proxies from univariate analyses

<table>
<thead>
<tr>
<th>Position (hg38)</th>
<th>LD (r²)</th>
<th>SNP (Ref&gt;Alt)</th>
<th>MAF in EUR</th>
<th>Functional Annotation</th>
<th>CR</th>
<th>Promoter histone marks</th>
<th>Enhancer histone marks</th>
<th>DNAse site</th>
<th>Proteins bound</th>
<th>eQTL</th>
<th>Motifs Changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr10:1688744</td>
<td>---</td>
<td>rs77027562 (A&gt;G)</td>
<td>0.02</td>
<td>Intronic</td>
<td>No</td>
<td>BRN</td>
<td>BLD</td>
<td></td>
<td></td>
<td></td>
<td>ERalpha-a, RXRA, Zfp281</td>
</tr>
<tr>
<td>Chr10:1675149</td>
<td>0.94</td>
<td>rs12258319 (G&gt;T)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>BLD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pax-5</td>
</tr>
<tr>
<td>Chr10:1675875</td>
<td>0.94</td>
<td>rs7077743 (C&gt;T)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>BLD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BDPI, CAC-binding-protein, HNF4, p300</td>
</tr>
<tr>
<td>Chr10:1676470</td>
<td>0.94</td>
<td>rs6560758 (T&gt;C)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>ESDR, IPSC, BLD, LNG</td>
<td>ESC, BLD</td>
<td>CTCF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr10:1678882</td>
<td>0.94</td>
<td>rs10751814 (A&gt;G)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>GR, Rad21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr10:1680294</td>
<td>0.94</td>
<td>rs7089727 (A&gt;G)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>ESDR, IPSC, BLD, LNG</td>
<td>ESC, BLD</td>
<td>CTCF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr10:1681695</td>
<td>0.94</td>
<td>rs6560759 (T&gt;C)</td>
<td>0.98</td>
<td>Intronic</td>
<td>No</td>
<td>ESDR, BLD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chr10:1687566</td>
<td>0.94</td>
<td>rs79784382 (T&gt;A)</td>
<td>0.02</td>
<td>Intronic</td>
<td>No</td>
<td>BRN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cphx, Duxl, HNF6, Hmx, Hoxa13, Pbx1, Pbx3</td>
</tr>
</tbody>
</table>

Abbreviations: LD, linkage disequilibrium; MAF, minor allele frequency; EUR, European; CR, conserved region; eQTL, expression quantitative trait loci. Tissue groups: BRN, brain cells; BLD, blood and T-cells; ESDR, embryonic stem cell derived cells; IPSC, induced pluripotent stem cells; LNG, lung cell; ESC, embryonic stem cells. Proxies were defined as variants in LD (r²>0.8) with the index SNP rs77027562 (bolded) in 1000 genomes project Phase 1 data for Europeans. All data was accessed using HaploReg v4.1 available at: http://www.broadinstitute.org/mammals/haploreg/documentation_v4.1.html. Both conservation prediction algorithms, GERP and SiPhy-omega, were used. Only eQTLs for ADAR genes (5 genes) are given.

Table 4: Association between RNA editing genes and EOC susceptibility.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAD1</td>
<td>210 (210)</td>
<td>98</td>
<td>112 (MAF&lt;0.01)</td>
<td>0.698</td>
<td>0.749</td>
<td>0.300</td>
<td>0.433</td>
<td>0.054</td>
<td>0.179</td>
<td>0.857</td>
<td>0.804</td>
<td>0.696</td>
<td>0.635</td>
</tr>
<tr>
<td>ADAR1</td>
<td>155 (155)</td>
<td>50</td>
<td>105 (MAF≥0.01)</td>
<td>0.020</td>
<td>0.022</td>
<td>0.110</td>
<td>0.101</td>
<td>0.005</td>
<td>0.008</td>
<td>0.943</td>
<td>0.841</td>
<td>0.780</td>
<td>0.411</td>
</tr>
<tr>
<td>ADAR2</td>
<td>754 (754)</td>
<td>301</td>
<td>4563 (MAF&lt;0.01)</td>
<td>0.861</td>
<td>0.894</td>
<td>0.623</td>
<td>0.720</td>
<td>0.134</td>
<td>0.496</td>
<td>0.338</td>
<td>0.208</td>
<td>0.105</td>
<td>0.041</td>
</tr>
<tr>
<td>ADAR3</td>
<td>2656 (2654)</td>
<td>787</td>
<td>1867 (MAF≥0.01)</td>
<td>0.216</td>
<td>0.266</td>
<td>0.334</td>
<td>0.541</td>
<td>0.587</td>
<td>0.470</td>
<td>0.002</td>
<td>0.031</td>
<td>0.234</td>
<td>0.502</td>
</tr>
<tr>
<td>SND1</td>
<td>1528 (1527)</td>
<td>764</td>
<td>763 (MAF&lt;0.01)</td>
<td>0.630</td>
<td>0.809</td>
<td>0.703</td>
<td>0.376</td>
<td>0.919</td>
<td>0.895</td>
<td>0.632</td>
<td>0.204</td>
<td>0.773</td>
<td>0.535</td>
</tr>
</tbody>
</table>
selection. Efforts to replicate these findings are needed; data will be available soon from a large, independent cohort of EOC cases genotyped by OCAC for this purpose (Amos et al, The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers (provisionally accepted, *CEBP*). Mechanistic studies to reveal how ADAR polymorphisms may affect oncogenic phenotypes will also be required, as will systematic investigations of the genomic landscape and clinical relevance of RNA editing in EOC using data from TCGA or other sources.

In summary, this study provides data to support the hypothesis that germline polymorphisms in ADAR related genes may influence gene expression and susceptibility to EOC. Further investigations are needed to determine whether inherited and acquired alterations affecting RNA editing serve as biological mechanisms to promote the development of EOC.

**MATERIALS AND METHODS**

**Study population**

A total of 41 studies (32 case-control and 9 case-only) from OCAC contributed to this investigation (Supplementary Table 1). Briefly, cases were women diagnosed with histologically confirmed primary invasive EOC (95%), fallopian tube cancer (1%), or primary peritoneal cancer (4%). Controls were women without cancer and with at least one intact ovary on the reference date. Individual studies were grouped into 26 case-control strata. All studies provided data on disease status, age at diagnosis/interview, self-reported racial group, and histologic subtype.

**Genotyping, quality control (QC), and imputation**

Peripheral blood was the primary source of germline DNA and was collected in the course of clinical care or research at each of the participating sites. The candidate SNPs selected for the current investigation were genotyped using a custom Illumina Infinium iSelect Array as part of the international Collaborative Oncological Gene-environment Study (iCOGS), an effort to evaluate 211,155 genetic variants for association with cancer risk [17].

Briefly, OCAC genotyping was conducted at McGill University and Génome Québec Innovation Centre (Montréal, Canada) and Mayo Clinic Medical Genomics Facility. Each 96-well plate well contained 250ng genomic DNA (or 500 ng whole genome-amplified DNA). Raw intensity data files were sent to the COGS data coordination center at the University of Cambridge for genotype calling and QC using the GenCall algorithm.

![Figure 3: Box-plot showing that ADAR1 tumor tissue expression differed, albeit only slightly, by rs1127313 genotype (p = 0.027).](image-url)
Sample and SNP quality control procedures have been described previously; in brief, samples were excluded with call rates < 95%, >1% discordance, < 80% European ancestry, or ambiguous gender, and SNPs were excluded with call rates < 95% or monomorphism [18, 19].

To improve genomic coverage and power [14], we imputed genotypes based on data from the 1000 Genomes Project (1KGP); we used IMPUTE2 version 2 after pre-phasing with SHAPEIT [20]. All 14 populations in the 1KGP were used as the reference. Before imputation, we excluded poorly performing SNPs according to the genotyping success rates, deviation from Hardy-Weinberg equilibrium (HWE) ($P < 1\times10^{-7}$), and replicate errors. To ensure the quality of the imputed genotypes, maximum likelihood genotype imputation was carried out and an estimate of the squared correlation between the imputed and true genotypes was calculated. Imputation quality is significantly decreased for low and rare frequency variants [21]. To be more inclusive of rare variants, we considered imputed SNPs with an $r^2 > 0.40$ as well-imputed [22] and included them in our analyses. The average imputation quality for included variants is detailed in Supplementary Table 4, overall and by MAF categories.

Gene and SNP selection

Five candidate genes were chosen for this study based on published literature which directly showed or suggested roles in the regulation of A-to-I RNA editing [1, 4, 23]. The genes included adenosine deaminase domain containing 1 (ADAD1), adenosine deaminase, RNA-specific (ADAR/ADAR1), adenosine deaminase, RNA-specific, B1 (ADARB1/ADAR2), adenosine deaminase, RNA-specific, B2 (ADAR3/ADARB2), and staphylococcal nuclease and Tudor domain containing 1 (SNDI). In total, 5,303 SNPs in the 5 genes, 77 genotyped directly and 5,226 imputed, were available for statistical analysis.

Population stratification

HapMap DNA samples from European (CEU, $n = 60$), African (YRI, $n = 53$) and Asian (JPT+CHB, $n = 88$) populations were also genotyped as part of the same custom Illumina iSelect Array. The program LAMP [24] was used to estimate intercontinental ancestry based on the HapMap (release no. 23) genotype frequency data for these three populations. Eligible subjects with greater than 90 percent European ancestry were defined as European ($n = 39,773$). We then used a set of 37,000 unlinked autosomal markers to perform principal components analysis within each major population subgroup. To enable this analysis on very large sample sizes we used an in-house program written in C++ using the Intel MKL libraries for eigenvectors (available at http://ccge.medschl.cam.ac.uk/software/).

**Statistical analysis**

Descriptive statistics were calculated in terms of means and standard deviations for continuous variables and frequencies and percents for categorical variables. The primary association analysis focused on individuals of European ancestry. Unconditional logistic regression was used to estimate odds ratios (OR) and their 95% confidence intervals (CI) between genotype and case status under a log-additive genetic model, with adjustment for the first five principal components representing sub-European ancestry. Due to the heterogeneous nature of EOC, subgroup analyses were conducted to estimate genotype-specific odds ratios by histologic subtype: serous, endometrioid, mucinous, and clear cell carcinomas. False discovery rates (FDR) [25] were used to adjust for multiple comparisons, and FDR of 15% was used to declare significance.

Two methods of gene-level evaluations were also conducted to combine association evidence from SNPs within each gene evaluated: the Admixture Maximum Likelihood (AML) Test [26] and the Sequence-Kernel Association test for the combined effect of common and rare variants (SKAT-CR) [27]. AML is an approach that simultaneously examines the global null hypothesis (of no SNP-outcome associations) and estimates the proportion of underlying false hypotheses. The AML uses univariate SNP-level results to calculate the AML Cochran-Armitage Trend test. Compared to other methods, AML has been shown to have similar or higher statistical power to detect associations except under the unlikely scenario that greater than 20% of all variants are associated with the outcome [26]. SKAT-CR evaluates the cumulative effect of rare and common variants, but does not consider low-frequency variants. These gene-level approaches were undertaken to complement SNP-level findings, and aimed to reduce the degrees of freedom, avoid model-fitting issues due to multicollinearity from LD, and to improve statistical power. The Bonferroni method was used to account for multiple comparisons.

Expression quantitative trait locus (eQTL) analysis was performed to examine for association between genotype ($n = 5,303$, imputed as above in $n = 5$ genes) and corresponding gene expression for the 5 candidate RNA editing genes. Matched genotype and gene expression profiling data were obtained for 402 high-grade serous EOC samples evaluated in the Cancer Genome Atlas (TCGA) Project using previously described methods [19]. Briefly, germline genotypes and matched tumor gene expression data were downloaded from the TCGA data portal. To conduct the eQTL analysis, we used germline genotypes of SNPs/proxies as independent variables and expression levels as traits. Expression levels between minor allele carriers versus non-carriers were compared using the Wilcoxon rank sum statistic. Haploreg v4.1 http://www.broadinstitute.org/mammals/haploreg/
haploreg.php) [28] was used to evaluate the putative function of candidate SNPs.

**CONFLICTS OF INTEREST**

None

**GRANT SUPPORT**

Funding for this study was supported in part by the US National Institute of Health (R01-CA-11343 and R01-CA114343-S1) 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; BAV: ELAN funds of the University of Erlangen-Nuremberg. BEL: Nationaal Kankerplan, we would like to thank Gilian Peuteman, Thomas Van Brussel, Annick Van den Broeck and Joke De Roover for technical assistance; DOV: U.S. National Institutes of Health R01-CA112523 and R01-CA87538; GER: German Federal Ministry of Education and Research, Programme of Clinical Biomedical Research (01 GB 9401) and the German Cancer Research Center (DKFZ). GRR: Roswell Park Cancer Institute Alliance Foundation, P30 CA016056. HAW: U.S. National Institutes of Health (R01-CA58598, N01-CN-55424 and N01-PC-67001); HJO and HMO: Intramural funding; Rudolf-Bartling Foundation; HOC: Helsinki University Research Fund; 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; LAX: American Cancer Society Early Detection Professorship (SIOP-06-258-01-COUN) and the National Center for Advancing Translational Sciences (NCATS), Grant UL1TR000124; 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.; MAY: National Institutes of Health (R01-CA122443, P30-CA15083, P50-CA136393), Mayo Foundation; Minnesota Ovarian Cancer Alliance; Fred C. and Katherine B. Andersen Foundation; MCC: MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.; NEC: National Institutes of Health R01-CA54419 and P50-CA105009 and Department of Defense W81XWH-10-1-02802; NHS: National Institute of Health (UM1-CA176726, and R01-CA67262). The NHS would like to thank the participants and staff for their valuable contributions as well as all the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY; NJO: National Cancer Institute (NIH-K07 CA095666, R01-CA83918, NIH-K22-CA138563, and P30-CA072720) and the Cancer Institute of New Jersey, and NCI CCSG award (P30-CA008748); NOR: Helse Vest, The Norwegian Cancer Society, The Research Council of Norway; NTH: Radboud University Medical Centre; ORE: OHSU Foundation; OVA: This work was supported by Canadian Institutes of Health Research grant (MOP-86727) and by NIH/NCI 1 R01CA160669-01A1; POC: Pomeranian Medical University; POL: Intramural Research Program of the NCI; PVD: Herlev Hospitals Forskningsråd, Direktor Jacob Madsens og Hustru Olga Madsens fond, Arvid Nilssons fond, Gangsted fonden, Herlev Hospitals Forskningsråd and Danish Cancer Society; RMH: Cancer Research UK (no grant number is available); 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 Lucecarini, Caroline Baynes, Don Conroy; SRO: Cancer Research UK (C536/A13086, C536/A6689) and Imperial Experimental Cancer Research Centre (C1312/A15589) and to thank all members of Scottish Gynaecological Clinical Trials group and COTROC1 investigators; STA: US National Institutes of Health U01-CA71966, R01-CA16056, K07-CA143047, and U01-CA69417 for recruitment of controls by the Cancer Prevention Institute of California; TOR: NIH grants R01 CA63678, CA63682, and CA149429; UCI: NIH R01-CA058860, and the Lon V Smith Foundation grant VLS-39420; 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;
REFERENCES


