Tagging Single Nucleotide Polymorphisms in Cell Cycle Control Genes and Susceptibility to Invasive Epithelial Ovarian Cancer


Departments of Medicine, Stanford, California; University of Copenhagen, Copenhagen, Denmark; Aarhus University Hospital, Skejby, Aarhus, Denmark; Roswell Park Cancer Institute, Buffalo, New York; The Queensland Institute of Medical Research, Post Office Royal Brisbane Hospital, Australia; Cancer Research Center, University of Hawaii, Honolulu, Hawaii; Department of Epidemiology and University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania; Mayo Clinic College of Medicine, Rochester, Minnesota; H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; University of Southern California, Keck School of Medicine, Los Angeles, California; Duke University Medical Center, Durham, North Carolina; Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Institute of Oncology and Cancer Center, Warsaw, Poland; Nofer Institute of Occupational Medicine, Lodz, Poland; and Peter MacCallum Cancer Institute, Melbourne, Australia

Abstract

High-risk susceptibility genes explain <40% of the excess risk of familial ovarian cancer. Therefore, other ovarian cancer susceptibility genes are likely to exist. We have used a single nucleotide polymorphism (SNP)–tagging approach to evaluate common variants in 13 genes involved in cell cycle control—CCND1, CCND2, CCND3, CCNE1, CDK2, CDK4, CDK6, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C, and CDKN2D—and risk of invasive epithelial ovarian cancer. We used a two-stage, multicenter, case-control study. In stage 1, 88 SNPs that tag common variation in these genes were genotyped in three studies from the United Kingdom, United States, and Denmark (~1,500 cases and 2,500 controls). Genotype frequencies in cases and controls were compared using logistic regression. In stage 2, eight other studies from Australia, Poland, and the United States (~2,000 cases and ~3,200 controls) were genotyped for the five most significant SNPs from stage 1. No SNP was significant in the stage 2 data alone. Using the combined stages 1 and 2 data set, CDKN2A rs3731257 and CDKN1B rs2066827 were associated with disease risk (unadjusted \( P \) trend = 0.008 and 0.036, respectively), but these were not significant after adjusting for multiple testing. Carrying the minor allele of these SNPs was found to be associated with reduced risk [OR, 0.91 (0.85–0.98) for rs3731257; and OR, 0.93 (0.87–0.995) for rs2066827]. In conclusion, we have found evidence that a single tagged SNP in both the CDKN2A and CDKN1B genes may be associated with reduced ovarian cancer risk. This study highlights the need for multicenter collaborations for genetic association studies.

Introduction

Several genes have been identified that confer high-penetration susceptibility to epithelial ovarian cancer. The main protagonists are \( BRCA1 \) and \( BRCA2 \). These genes are responsible for about half of all families containing two or more ovarian cancer cases in first-degree relatives and most families containing several ovarian cancer cases (≥3) or in which ovarian and breast cancers occur together (1, 2).

However, the known susceptibility genes explain <40% of the excess familial ovarian cancer risk (3). If other highly penetrant ovarian cancer susceptibility genes exist, they are likely to be rare. One possible explanation for the residual risks is the existence of several common but only moderately or low-penetrance susceptibility alleles in the population (4).

There have been several studies that have attempted to identify common variants in genes that may be associated with increased risk of ovarian cancer. Candidate genes have usually been selected based on biological plausibility. These include genes in pathways controlling steroid hormone metabolism (e.g., progesterone receptor, androgen receptor, \( CYP17 \), prohibitin; refs. 5–14), DNA repair

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Simon Gayther, Translational Research Laboratories, Windeyer Institute, University College London, 46 Cleveland Street, London W1T 4JF, United Kingdom. Phone: 44-20-7679-9204; Fax: 44-20-7679-9687; E-mail: s.gayther@ucl.ac.uk.

©2007 American Association for Cancer Research. doi:10.1158/0008-5472.CAN-06-3261

www.aacrjournals.org
BRCA1, BRCA2, RAD51, MSH2; refs. 15–17) and in putative oncogenes and tumor suppressor genes (TP53, RB1, HRAS1, STK15; refs. 18–23). Thus far, reported positive associations include an increased ovarian cancer risk for two PGR haplotypes (8); a protective effect for the PGR promoter +331A allele in endometrioid ovarian tumors (9); an increased risk of borderline ovarian cancer associated with the Pro72Arg polymorphism in the TP53 gene (20); and an increased risk associated with the polymorphism in the STK15 oncogene (23). However, none of these associations are conclusive due to the small size of the initial studies, and there are no reports indicating that they have been validated in additional populations.

There are no published reports describing the association between variants in cell-cycle control genes and ovarian cancer susceptibility. In general, cancer is associated with a breakdown in the mechanisms that regulate cell division. In mammalian cells, cell division is controlled by the activity of cyclin-dependent kinases (CDK) and their essential activating coenzymes, the cyclins and CDK inhibitors (reviewed in refs. 24, 25). CDK activity is regulated on several levels, including cyclin synthesis and degradation, phosphorylation, and dephosphorylation. The interaction between CDKs and cyclins is tightly controlled to ensure an ordered progression through the cell cycle from G1 to S to G2. The events occurring before DNA synthesis are probably the most thoroughly understood. The key event in cell cycle regulation is transversion of the restriction (R) point late in G1, which is crucial to the cell's destiny toward division, differentiation, senescence, or apoptosis. It is believed that once the R point has been overcome, cell cycle progression occurs almost automatically. In the event of cellular stress, there are several proteins that can inhibit the cell cycle in G1.

### Table 1A. Characteristics of the 11 ovarian cancer case control studies (listed alphabetically) included in the Ovarian Cancer Association Consortium pooled study

<table>
<thead>
<tr>
<th>Study name</th>
<th>Location</th>
<th>Total number of cases*</th>
<th>Total number of controls*</th>
<th>Age range (cases)</th>
<th>Age range (controls)</th>
<th>Reference</th>
<th>DNA source</th>
<th>Days from diagnosis to blood draw</th>
<th>Number of SNPs genotyped</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOCs and ACS (ovarian cancer)</td>
<td>Multiregional, Australia</td>
<td>802 (71)</td>
<td>933 (42)</td>
<td>23–80</td>
<td>19–81</td>
<td>9</td>
<td>Blood lymphocytes</td>
<td>0–1,663 (103)</td>
<td>5</td>
</tr>
<tr>
<td>FROC (Family Registry for Ovarian Cancer Study)</td>
<td>Stanford, CA</td>
<td>327 (39)</td>
<td>429 (59)</td>
<td>23–64</td>
<td>19–66</td>
<td>16, 17, 21, 23</td>
<td>Blood lymphocytes; mouthwash</td>
<td>55–1,477 (281)</td>
<td>88</td>
</tr>
<tr>
<td>Hawaii Ovarian Cancer Study</td>
<td>Hawaii</td>
<td>313 (241)</td>
<td>574 (426)</td>
<td>18–87</td>
<td>19–88</td>
<td>33</td>
<td>Blood lymphocytes; mouthwash</td>
<td>16–2,026 (375)</td>
<td>5</td>
</tr>
<tr>
<td>HOPE (hormones and ovarian cancer prediction study)</td>
<td>Pittsburgh</td>
<td>57 (1)</td>
<td>152 (8)</td>
<td>34–80</td>
<td>44–79</td>
<td>–</td>
<td>Blood lymphocytes; mouthwash</td>
<td>0–280 (149)</td>
<td>5</td>
</tr>
<tr>
<td>LAC-CCOC</td>
<td>Los Angeles, CA</td>
<td>659 (211)</td>
<td>819 (198)</td>
<td>18–84</td>
<td>21–78</td>
<td>8</td>
<td>Blood lymphocytes</td>
<td>46–1,340 (252)</td>
<td>5</td>
</tr>
<tr>
<td>MALOVA (Malignant Ovarian Cancer Study)</td>
<td>Copenhagen, Denmark</td>
<td>446</td>
<td>1221 (0)</td>
<td>32–80</td>
<td>35–79</td>
<td>16, 17, 21, 23</td>
<td>Blood lymphocytes; Blood lymphocytes</td>
<td>All preoperative</td>
<td>88</td>
</tr>
<tr>
<td>Mayo Clinic Rochester, ovarian cancer case-control study</td>
<td>Minnesota</td>
<td>317 (13)</td>
<td>462 (35)</td>
<td>28–91</td>
<td>23–90</td>
<td>32</td>
<td>Blood lymphocytes</td>
<td>0–365 (31)</td>
<td>5</td>
</tr>
<tr>
<td>NCOCs</td>
<td>North Carolina</td>
<td>610 (91)</td>
<td>843 (156)</td>
<td>22–74</td>
<td>20–75</td>
<td>9</td>
<td>Blood lymphocytes</td>
<td>29–1,259 (187)</td>
<td>5</td>
</tr>
<tr>
<td>POCS</td>
<td>Warsaw and Lodz</td>
<td>264</td>
<td>625 (0)</td>
<td>24–74</td>
<td>24–74</td>
<td>–</td>
<td>Blood lymphocytes; mouthwash</td>
<td>0–525 (82)</td>
<td>5</td>
</tr>
<tr>
<td>SEARCH ovarian cancer</td>
<td>Cambridge, United Kingdom</td>
<td>731 (42)</td>
<td>855 (4)</td>
<td>21–74</td>
<td>39–77</td>
<td>16, 17, 21, 23</td>
<td>Blood lymphocytes</td>
<td>140–4,850 (1,231)</td>
<td>88</td>
</tr>
</tbody>
</table>

*Numbers of subjects that are non-white are given in parentheses.
†The mean is given in parentheses.
For example, in response to DNA damage, p53 accumulates in the cell and induces p21/Cdkn1a-mediated inhibition of cyclin D/CDK.

Somatic alterations of genes involved in the G1 phase of the cell cycle, including the cyclins, CDKs, and CDK inhibitors, are common events in neoplastic development for multiple tumor types; but different cell cycle proteins seem to be targets for different cancers. A reason for this could be that although the basic mechanisms of proliferative control are identical in different tissues, the maintenance of normal cell-cycle control is also partly regulated in a tissue-specific manner. For ovarian cancer, overexpression of cyclin D1 (CCND1) has been reported in most borderline and invasive ovarian cancers (26, 27); cyclin D2 was found to be overexpressed in about 40% of ovarian tumors compared with normal ovarian tissues (28); P16INK4a (CDKN2A) is relatively frequently lost due to deletion, loss of expression, and hypermethylation in ovarian cancer cell lines and primary tumors (29, 30); and in some primary ovarian cancers, homozygous deletions of P16INK4a involving the neighboring P15INK4b (CDKN2B) gene are associated with a poor prognosis for ovarian cancer cases (31). Together, these data suggest an important role for CDKs, cyclins, and CDK inhibitors in ovarian cancer development.

The aim of this study was to evaluate the association between common variants in 13 genes coding for cyclins, CDKs, and CDK inhibitors and susceptibility to epithelial ovarian cancer. We used a two-stage case-control study design in which single nucleotide polymorphisms (SNP) that tag all known common variants in these genes were first genotyped in three studies. The most significant hits were then genotyped in eight additional ovarian cancer case-control studies from the international ovarian cancer association consortium (OCAC).

<table>
<thead>
<tr>
<th>Study</th>
<th>Case ascertainment</th>
<th>Control ascertainment*</th>
<th>Participation rates †</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOCs and ACS Australia</td>
<td>Diagnosed from 2002 onward; recruited through surgical treatment centers throughout Australia and cancer registries of Queensland South Australia and West Australia (AOCs) and cancer registries of New South Wales and Victoria (ACS)</td>
<td>Randomly selected from Commonwealth electoral roll. Frequency matched to cases for geographic region</td>
<td>Cases: 84%; controls: 47%</td>
</tr>
<tr>
<td>FROC, USA</td>
<td>Consecutive cases diagnosed from 1997–2002 in Greater Bay Area Cancer Registry San Francisco</td>
<td>Random-digit dial identification from study area</td>
<td>Cases: 75%; controls: 75%</td>
</tr>
<tr>
<td>Hawaii, USA</td>
<td>Rapid case ascertainment through Hawaii Tumor Registry from 1993 onward</td>
<td>Randomly selected from Hawaii Department of Health Annual Survey of the representative households</td>
<td>Cases: 58.1%; controls: 62.8%</td>
</tr>
<tr>
<td>HOPE, USA</td>
<td>Variable source including physician offices cancer registries and pathology databases from counties of western Pennsylvania, eastern Ohio, and western New York</td>
<td>Identified in same regions. All participants undergo home interviews</td>
<td>Cases: 69%; controls: 81%</td>
</tr>
<tr>
<td>LAC-CCOC, USA</td>
<td>Rapid case ascertainment through Los Angeles Cancer Surveillance program from 1993 onward</td>
<td>Neighborhood-recruited controls</td>
<td>Cases: 73%; controls: 73%</td>
</tr>
<tr>
<td>MALOVA, Denmark</td>
<td>Incident cases (35–79 yrs) diagnosed 1994–1999 from municipalities of Copenhagen and Frederiksberg and surrounding counties</td>
<td>Random sample of general female population (35–79 yrs) in study area selected using computerized Central Population Register</td>
<td>Cases: 79%; controls: 67%</td>
</tr>
<tr>
<td>Mayo, USA</td>
<td>Cases attending Mayo Clinic diagnosed from 2000 onward identified in a six-state surrounding region</td>
<td>Identified through Mayo Clinic; healthy women seeking general medical examination</td>
<td>Cases: 84%; controls: 65%</td>
</tr>
<tr>
<td>NCOCS, USA</td>
<td>Cases from 1999 onward identified from 48 counties within the region by rapid-case ascertainment</td>
<td>Controls identified from the same region as cases</td>
<td>Cases: 70%; controls: 63%</td>
</tr>
<tr>
<td>POCS, Poland</td>
<td>Cases collected from cities of Warsaw and Lodz 2001–2003 by rapid ascertainment at participating hospitals</td>
<td>Identified at random through the Polish electronic system; stratified by city and 5-yr age categories</td>
<td>Cases: 71%; controls: 67%</td>
</tr>
<tr>
<td>SEARCH, UK</td>
<td>Cases &lt;70 yrs from East Anglian West Midlands and Trent regions of England; prevalent cases diagnosed 1991–1998; incident cases diagnosed 1998 onward</td>
<td>Selected from the EPIC-Norfolk cohort of 25,000 individuals aged 45–74 based in the same geographic regions as the cases</td>
<td>Cases: 67%; controls: 84%</td>
</tr>
</tbody>
</table>

*All controls were frequency matched to cases for age and ethnicity.
†Participation rates are the percentage of those taking part as proportion of those invited to take part.
Table 2. Details of tSNP selection for cell cycle control genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Size (kb)</th>
<th>Data</th>
<th>% of gene resequenced by EGP</th>
<th>No. common variants</th>
<th>No. tSNPs</th>
<th>Additional SNPs genotyped</th>
<th>Additional mean $r^2$ Number with $r^2 &gt; 0.8$</th>
<th>$r^2$ for poorly tagged SNPs</th>
<th>LD blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCND1</td>
<td>16.0</td>
<td>EGP</td>
<td>100</td>
<td>12 $^\dagger$</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>0.82</td>
<td>10</td>
</tr>
<tr>
<td>CCND2</td>
<td>34.0</td>
<td>EGP</td>
<td>93</td>
<td>24</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>0.89</td>
<td>21</td>
</tr>
<tr>
<td>CCND3</td>
<td>9.0</td>
<td>EGP</td>
<td>88</td>
<td>13</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0.92</td>
<td>12</td>
</tr>
<tr>
<td>CCNE1</td>
<td>13.5</td>
<td>EGP</td>
<td>84</td>
<td>26</td>
<td>4</td>
<td>11</td>
<td>0</td>
<td>0.57</td>
<td>6</td>
</tr>
<tr>
<td>CDK2</td>
<td>7.6</td>
<td>EGP</td>
<td>98</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0.54</td>
<td>2</td>
</tr>
<tr>
<td>CDK4</td>
<td>6.7</td>
<td>EGP</td>
<td>96</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.86</td>
<td>4</td>
</tr>
<tr>
<td>CDK6</td>
<td>220.8</td>
<td>HapMap</td>
<td>18</td>
<td>41</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0.96</td>
<td>41</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>10.9</td>
<td>EGP</td>
<td>85</td>
<td>27</td>
<td>10 $^\dagger$</td>
<td>7</td>
<td>1</td>
<td>0.85</td>
<td>20</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>5.7</td>
<td>HapMap</td>
<td>100</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0.90</td>
<td>6</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>28.8</td>
<td>EGP</td>
<td>81</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0.97</td>
<td>12</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>13.3</td>
<td>EGP</td>
<td>94</td>
<td>20</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0.96</td>
<td>20</td>
</tr>
<tr>
<td>CDKN2C</td>
<td>8.9</td>
<td>EGP/HapMap</td>
<td>87</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>3</td>
</tr>
<tr>
<td>CDKN2D</td>
<td>4.1</td>
<td>EGP</td>
<td>63</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1.00</td>
<td>4</td>
</tr>
</tbody>
</table>

Abbreviation: N/A, not applicable.

*Additional SNPs that were genotyped that were not common in the data set used for tagging. These SNPs do not contribute to the estimates of tagging efficiency.

$^\dagger$One common SNP in EGP (environmental genome project; rs3212867) was monomorphic in the CEU HapMap population.

$^\ddagger$Assay for rs3176359 was too rare in our populations (MAF = 0.002).

Materials and Methods

**Study subjects.** In total, 11 different ovarian cancer case-control studies (summarized in Table 1A and 1B) contributed data to this study, three for stage 1 and eight for stage 2 (see also refs. 8–10, 16, 17, 21, 23, 32, 33 for further details). Most subjects were white Europeans in origin. Ten studies used population-based ascertainment for cases and controls, and one study was clinic based. All studies have received ethical committee approval, and all study subjects provided informed consent.

**Tag SNP selection.** A set of tagging SNPs (tSNP) was identified for 13 genes comprising four cyclins (CCND1, CCND2, CCND3, CCNE1); three CDKs (CDK2, CDK4, CDK6); six CDK inhibitors (CDKN1A, CDKN1B, CDKN2A, CDKN2B, CDKN2C, CDKN2D); the principal hypothesis underlying the study was that one or more common SNPs in these genes were associated with an altered risk of ovarian cancer. We therefore identified a set of tagging SNPs that efficiently tagged all the known common variants (minor allele frequency (MAF) $\geq 0.05$) and were likely to tag most of the unknown common variants.

The selection of tSNPs is most reliable where the gene has been resequenced in a sample of individuals sufficiently large to identify all common variants. The National Institute of Environmental Health Sciences (NIEHS) Environmental Genome Project (EGP) is currently resequencing candidate genes for cancer across a panel of 90/95 individuals representative of U.S. ethnicities. The original panel (P1-PDR90) of 90 individuals consists of 24 European Americans, 24 African Americans, 12 Mexican Americans, 6 Native Americans, and 24 Asian Americans, but the ethnic group identifiers are not available. Because it is known that there is greater genetic and haplotype diversity in individuals of African origin, we have identified and excluded 28 of the samples with the greatest African ancestry in this population by comparing the genotypes of the resequenced sample with genotypes for the same SNPs from the National Heart, Lung, and Blood Institute Variation Discovery Resource Project African American panel (http://pga.gs.washington.edu/finished_genes.html). Data from the remaining 62 individuals were used to identify tSNPs. Ideally, samples from Native American, Hispanic American, and Asian American individuals should also be removed; but there is less genetic diversity between these groups, and they cannot be excluded with any certainty.

When resequencing data were not available, we used data from the International HapMap Project (release 34/5 21-06-2005: last public release used in this study) to select tSNPs. The HapMap Project has genotyped a large number of SNPs in several populations, including 30 parent-offspring trios collected in 1980 from U.S. residents with northern and western European ancestry by the Centre d'Etude du Polymorphisme Humain (CEPH).

The best measure of the extent to which one SNP tags another SNP is the pairwise correlation coefficient ($r_{ps}$) because the loss in power incurred by using a marker SNP in place of a true causal SNP is directly related to this measure. We aimed to define a set of tSNPs such that all known common SNPs had an estimated $r_{ps} > 0.8$ with at least one tSNP. However, some SNPs are poorly correlated with other single SNPs but may be efficiently tagged by a haplotype defined by multiple SNPs, so-called “aggressive tagging,” thus reducing the number of tSNPs needed. As an alternative, therefore, we aimed for the correlation between each SNP and a haplotype of SNPs ($r_{hp}$) to be at $>0.8$. The program Tagger uses a strategy that combines the simplicity of pairwise methods with the potential efficiency of multimarker approaches (3). It begins by selecting a minimal set of markers such that all alleles to be captured are correlated at an $r_{ps} > 0.8$ with a marker in that set. Certain markers can be forced into the tag list or explicitly prohibited from being chosen as tags. After this, it tries to “peel back” the tag list by replacing certain tags with multimarker tests. Tagger avoids overfitting by only constructing multimarker tests from SNPs that are in strong linkage disequilibrium with each other, as measured by a pairwise LD score.

When an assay for one of the chosen tSNPs could not be designed, the SNP tagging process was repeated with that SNP excluded from the possible set of tSNPs. Table 2 shows details of the tSNP selection for each of the genes analyzed in this study.

---

21 http://www.broad.mit.edu/mpg/tagger/
Genotyping. Genotyping was carried out for 9 of the 11 studies using TaqMan (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Primers and probes were Assays-by-Design (Applied Biosystems). Following PCR amplification, end reactions were read on the ABI Prism 7900 using Sequence Detection Software (Applied Biosystems). Two studies [Australian Ovarian Cancer Study (AOCs) and Australian Cancer Study (ACS); Table 1A and B] used iPLEX technology (Sequenom) to perform genotyping. Failed genotypes were not repeated. Plates with genotype calling <95% were discarded. If there was genotyping discordance between duplicates within an experiment, the data were discarded and either repeated or the assay replaced with a different assay with similar tagging properties.

Each group carried out genotyping on their own samples. Therefore, we tested the quality of genotyping between laboratories for the five tSNPs analyzed by all groups using the HAPMAPPT01 panel of CEPH-Utah trios-standard plate supplied by Coriell. This panel includes 90 different DNA samples, five duplicate samples, and a negative template control in a 96-well plate format. We compared genotype call rates and concordance between studies. Call rates on these plates ranged from 96% to 99%. Overall concordance on these plates was >99%.

Statistical methods. Deviations of genotype frequencies in the controls from those expected under Hardy-Weinberg equilibrium (HWE) were assessed by $\chi^2$ tests [1 degree of freedom (d.f.)]. We treated the design as a single, two-stage multicenter study rather than hypothesis-generating and replication studies because joint analysis has been shown to be more efficient than replication analysis in a staged design (34). The primary tests of association were the univariate analyses between each tSNP and ovarian cancer. Genotype frequencies in cases and controls were compared for each study separately using $\chi^2$ tests for heterogeneity (2 d.f.). The data were then pooled, and a genetic model free test was carried out by comparing genotype frequencies in cases and controls using unconditional logistic regression with terms for genotype and study and an appropriate likelihood ratio test ($P_{het}$, 2 d.f.). We tested for heterogeneity between studies by comparing logistic regression models with and without a genotype-study interaction term using likelihood ratio tests. Genotypic specific risks with the common homozygote as the baseline comparator were estimated as odds ratios with associated 95% confidence limits by unconditional logistic regression. We also tested for rare allele dose effect (assuming a multiplicative genetic model) using $\chi^2$ tests for each study separately and unconditional logistic regression for the pooled data ($P_{het}$, 1 d.f.). Primary analyses were restricted to subjects of white European origin. Secondary analyses included all study subjects with additional adjustment for ethnic group.

In addition to the univariate analyses, we carried out specific haplotype tests for those combinations of alleles that tagged specific SNPs. We also...
carried out a general comparison of common haplotype frequencies in each gene haplotype block using the data from all the tSNPs in that block. Haplotype blocks were defined such that the common haplotypes (>5% frequency) accounted for at least 80% of the haplotype diversity. We considered haplotypes with >2% frequency in at least one study to be common. Rare haplotypes were pooled. For both specific haplotype marker tests and the general comparison of haplotype frequencies by haplotype block, haplotype frequencies and subject-specific expected haplotype frequencies (by haplotype block) and ovarian cancer was tested by comparing a model with terms for haplotype uncertainty given unphased genotype data (35). Subjects indicators were calculated separately for each study using the program TagSNPs. This implements an expectation substitution approach to account for haplotype uncertainty given unphased genotype data (35).

### Results

#### Identifying Tagging SNPs in Cell Cycle Control Genes

Using a combination of NIEHS resequencing data and data from the international HapMap project, we identified a total of 199 common SNPs (MAF > 0.05) in the 13 genes. One-hundred and sixty-seven SNPs were tagged with \( r^2 > 0.8 \), of which 10 were tagged by 5 different multimarker haplotypes. The remaining SNPs were tagged with \( r^2 \) between 0.07 and 0.74. These could not be tagged more efficiently because assays could not be designed for them. Eighty-one tSNPs were selected for this series. We also selected a further seven SNPs that were not found in the EGP or HapMap data sets used for tSNP selection. These SNPs do not contribute to the estimates of tagging efficiency. Thus, we identified 88 SNPs from 13 genes for genotyping. Eleven genes were tagged with an average \( r^2 > 0.8 \). For two genes (CDKN2A and CDKN2B), the best tagging we were able to achieve was \( r^2 = 0.54 \) and 0.57, respectively. These data are summarized in Table 2.

### Genotyping in Ovarian Cancer Cases and Controls

Genotyping was done in two stages. In stage 1, the 88 tSNPs described above were genotyped in three case-control populations from the United Kingdom (SEARCH), Denmark [Malignant Ovarian Cancer (MALOVA)], and the United States (Stanford). Combined, these studies comprise ~1,500 invasive ovarian cancer cases and 2,500 controls. In stage 2, we genotyped five SNPs from four genes that showed the strongest evidence of association in stage 1, in ~3,000 additional cases and 4,400 controls from another eight case-control studies.

#### Stage 1.

Nineteen out of 264 genotype distributions in controls deviated from HWE (\( P < 0.05 \)). In the majority of cases, deviations from HWE occurred in only one of the studies. The discrimination of genotypes for these assays was good. For rs3176319 (CDKN1A), the deviations from HWE were substantial in all three studies (\( P < 10^{-5} \)), and so this SNP was excluded from further analyses. rs3176359 in CDKN1A was also excluded from further analyses because it was found to be very rare in all three populations (MAF = 0.002).

We identified 28 SNPs with a \( P \leq 0.2 \), of which 9 SNPs had a \( P \leq 0.05 \) in tests for association. These data are summarized in Table 3. The complete data for all SNPs are given in Supplementary Tables S1 and S2. There was no evidence for association of genotype with age in controls, and as expected, age-adjusted risks were similar to unadjusted risks (data not shown). The haplotype frequencies and associated risks for the five multimarker haplotype tags are presented in Supplementary Table S3. There was no evidence for association at a significance level \( P \leq 0.05 \) for SNPs in 6 of the 13 genes (CDK2, CDK4, CCND1, CDKN1A, CDKN2C, CDKN2D).

Evidence of association at the \( P = 0.05 \) level was found for the following: five of seven SNPs in CCND1 (strongest association \( P \text{ trend} = 0.01 \)); one of seven SNPs in CCND3 (\( P=0.09 \)); 1 of 13 SNPs in CDK6 (\( P=0.04 \)); one of eight SNPs in CDKN1B (\( P=0.01 \)); and 1 of 17 SNPs in CDKN2A/2B (\( P=0.046 \)). Further details are provided in Table 3.

The global test of association with common haplotypes was not significant for any of the 13 genes (data not shown). One haplotype

### Table 4. Age-adjusted ovarian cancer risks associated with five tSNPs in cell cycle control genes for white cases and white controls only

<table>
<thead>
<tr>
<th>Gene</th>
<th>dbSNP</th>
<th>MAF</th>
<th>No. cases</th>
<th>No. controls</th>
<th>HetOR (95% CI)</th>
<th>HomOR (95% CI)</th>
<th>( P_{\text{adj}} )</th>
<th>( P_{\text{trend}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Combined analysis of all stage 2 studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td>rs7178</td>
<td>0.08</td>
<td>2,194</td>
<td>3,294</td>
<td>1.00 (0.86–1.18)</td>
<td>1.01 (0.95–1.06)</td>
<td>0.99</td>
<td>0.91</td>
</tr>
<tr>
<td>CCND1</td>
<td>rs603965</td>
<td>0.45</td>
<td>1,874</td>
<td>2,833</td>
<td>1.00 (0.87–1.15)</td>
<td>1.01 (0.86–1.20)</td>
<td>0.98</td>
<td>0.90</td>
</tr>
<tr>
<td>CDK6</td>
<td>rs8</td>
<td>0.21</td>
<td>2,188</td>
<td>3,298</td>
<td>1.02 (0.91–1.15)</td>
<td>0.91 (0.69–1.19)</td>
<td>0.69</td>
<td>0.83</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>rs2066827</td>
<td>0.23</td>
<td>2,202</td>
<td>3,297</td>
<td>1.05 (0.93–1.18)</td>
<td>0.88 (0.69–1.12)</td>
<td>0.36</td>
<td>0.82</td>
</tr>
<tr>
<td>CDKN2A/2B</td>
<td>rs3731257</td>
<td>0.27</td>
<td>2,190</td>
<td>3,290</td>
<td>0.92 (0.82–1.04)</td>
<td>0.89 (0.72–1.10)</td>
<td>0.29</td>
<td>0.13</td>
</tr>
<tr>
<td>B. Combined analysis for all studies, white subjects only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td>rs7178</td>
<td>0.08</td>
<td>3,607</td>
<td>5,725</td>
<td>1.09 (0.96–1.22)</td>
<td>1.12 (0.64–1.97)</td>
<td>0.37</td>
<td>0.12</td>
</tr>
<tr>
<td>CCND1</td>
<td>rs603965</td>
<td>0.45</td>
<td>3,285</td>
<td>5,236</td>
<td>1.02 (0.92–1.13)</td>
<td>1.13 (0.99–1.28)</td>
<td>0.15</td>
<td>0.084</td>
</tr>
<tr>
<td>CDK6</td>
<td>rs8</td>
<td>0.21</td>
<td>3,597</td>
<td>5,720</td>
<td>1.09 (1.00–1.19)</td>
<td>1.06 (0.86–1.31)</td>
<td>0.17</td>
<td>0.082</td>
</tr>
<tr>
<td>CDKN1B</td>
<td>rs2066827</td>
<td>0.25</td>
<td>3,618</td>
<td>5,719</td>
<td>0.98 (0.89–1.07)</td>
<td>0.79 (0.65–0.95)</td>
<td>0.044</td>
<td>0.036</td>
</tr>
<tr>
<td>CDKN2A/2B</td>
<td>rs3731257</td>
<td>0.27</td>
<td>3,601</td>
<td>5,705</td>
<td>0.89 (0.81–0.97)</td>
<td>0.87 (0.73–1.03)</td>
<td>0.021</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

22 http://locus.umdnj.edu/nigms/cgi/panel.cgi?id=2&query=HAPMAP01
that showed a significant frequency difference between cases and controls was observed in CDK6 block 3 (haplotype frequency, 0.2 in controls). This haplotype, comprising the rare allele of rs8 and common alleles of rs479049, rs445, and rs992519, showed increased risk for ovarian cancer compared with the most common haplotype in this block (OR, 1.19 [1.06–1.35]; \( P = 0.004 \)).

**Stage 2.** Five SNPs were genotyped in the stage 2 studies. These were rs7178 and rs60396 (both in CCND1), rs8 (CDK6), rs2066827 (CDKN1B), and rs3731257 (CDKNA2A-2B). The evidence for association for these SNPs from the stage 1 analyses (adjusted for age and ethnicity) ranged from \( P \) trend 0.003 to 0.046. There was some evidence for between-study heterogeneity for rs8 (\( P = 0.036 \)). None of the SNPs were significantly associated with ovarian cancer when considering the stage 2 studies as a replication set (Table 4A); rs3731257 was the most significant (\( P \) trend = 0.13).

**Combined analysis.** As joint analysis with appropriate adjustment for multiple testing has been shown to be more powerful than replication analysis (34), we also carried out a combined analysis of all the data. Figure 1 shows the genotype-specific risks (estimated as odds ratios) for each study and for the combined data for the analyses restricted to white subjects. Only two of the five SNPs were significant at the 0.05 level (Table 4B). These were rs2066827 (CDKN1B) and rs3731257 (CDKNA2A-2B; \( P \) trend = 0.036 and 0.008, respectively). The rare alleles for both SNPs were associated with reduced ovarian cancer risks [rs2066827 HetOR, 0.98 (0.89–1.07); HomOR, 0.79 (0.65–0.95); and rs3731257 HetOR, 0.89 (0.81–0.97); HomOR, 0.87 (0.73–1.03)]. There was little change in the association of rs2066827 when the data for all ethnic groups were included (\( P \) trend = 0.006); but for rs3731257, the association was almost completely attenuated (\( P = 0.16 \)). Standard methods for correcting for multiple testing, such as the Bonferroni correction, are too conservative when hypotheses are correlated, as is the case here. We have therefore used a permutation procedure modified to account for the staged design (36). A \( P \) value smaller than the most significant result was observed in 368 out of 1,000 permutations, which is equivalent to an adjusted \( P \) trend of 0.37.

Statistical power to identify subgroup effects from these studies is limited; but there were sufficient data to carry out an analysis with cases restricted to serous and nonserous histopathologic subtypes. In total, there were \( ~2,000 \) cases of serous histology in the 11 studies combined and 1,600 nonserous cases. The analysis based on this histologic subgroup stratification suggested marginal associations for rs3731257 with both serous (\( P \) trend = 0.030) and nonserous cancers (\( P \) trend = 0.042) and for rs7178 (\( P \) trend = 0.034), rs8 (\( P \) trend = 0.046), and rs2066827 (\( P \) trend = 0.023) with nonserous cancers only (Table 5).

**Discussion.** We have assessed the effect of tSNPs in 13 cell cycle control genes on the risks of invasive epithelial ovarian cancer. We chose to analyze genes involved in the control of cell division (cyclins, CDK, and CDK inhibitors) because several of these genes have previously been implicated in the somatic development of multiple tumor types. To our knowledge, this is the first report describing the association between these genes and ovarian cancer susceptibility.

A two-stage design was used for this study. In the first stage, 88 tSNPs were genotyped in three ovarian cancer case-control studies. The best five SNPs that were marginally associated with epithelial ovarian cancer were then genotyped in a further eight case-control studies from the international Ovarian Cancer Association Consortium, but none was significant in the second stage. However, the power to detect an effect in the replication studies was just 70% to detect an allele with frequency and effect size similar to CDKN1B rs2066827 (the most significant association from stage 1) at a type I
error rate of 0.01. Because a combined analysis with adjustment for multiple testing has been shown to have greater power than a replication analysis, we also carried out a joint analysis of the stages 1 and 2 data. Two SNPs, CDKN1B-rs2066827 and CDKN2A/2B-rs3731257, showed weak evidence for association, but neither was highly statistically significant (naive $P_{\text{trend}} = 0.036$ and 0.008, respectively), but these were not significant after adjustment for multiple testing using a permutation procedure that allows for the correlation between the SNPs. As an alternative to correcting for multiple testing, some authors have suggested that simple but stringent criteria should be applied to statistical tests for genetic association; e.g., $P < 10^{-4}$ for candidate gene studies or even $P < 10^{-7}$ for genomewide significance (37). Using these criteria, it can be estimated that 26,448 cases and an equal number of controls would be needed to confirm the effect for rs2066827 to detect a codominant allele with a minor allele frequency of 0.25 that confers a risk of 1.08 with 90% power at a type I error rate of $10^{-4}$.

Because the selection of SNPs in this study was based on a tagging approach rather than on putative function, we are unable to comment in great detail on the possible functional significance of these findings. The rs2066827 variant in CDKN1B is potentially functional. It encodes a nonsynonymous amino acid change Val$^{109}$Gly in CDKN1B and is situated in the interaction surface of CDK1B and its negative regulator p38$^{\text{a,b}}$, in the region spanning amino acid residues 97 to 151. This SNP could therefore alter the interaction between CDK1B and p38$^{\text{a,b}}$. The Val$^{109}$Gly variant has previously been reported in association with advanced prostate cancer risk (38).

The rs3731257 SNP in CDKN2A/2B is of no obvious functional significance, but it could be in linkage disequilibrium with another functional SNP within the gene. CDKN2A is a G1 CDK inhibitor that binds to CDK4 and CDK6 and prevents their association with D-type cyclins, thereby facilitating CDK4/6–cyclin D–mediated phosphorylation and inactivation of retinoblastoma protein and entry into S-phase. CDKN2B inhibits MDM2-mediated degradation of P53; thus, its loss would lead to a reduction in levels of the P53 protein (39, 40). The known function of both proteins suggests a plausible role in tumor development.

An alternative explanation for a false positive association is hidden population stratification. This occurs when allele frequencies differ between population subgroups, and cases and controls are drawn differentially from those subgroups. However, this study comprised 11 different case-control populations from England, Denmark, Poland, Australia, and several states throughout the United States. It is unlikely that population stratification is important because any population stratification will be study specific. Furthermore, the combined study was of sufficient size to allow the analyses to be restricted to white-only cases and controls, which represented the vast majority of all subjects.

Disease heterogeneity could also lead to false-positive reporting or mask the presence of true associations. In this study, we stratified cases according to histologic subtype and did the analysis for the combined data after dividing cases into serous and nonserous subgroups. Inevitably, this leads to a loss in statistical power, and although some marginal associations were found, these data should be treated with caution. Cases were collected from multiples of different centers, and there are likely to be variations between studies in the completeness of these data and in the reporting of different histopathologic subtypes.

There was no evidence of association with disease risk for polymorphisms in CCND2, CCNE1, CDK2, CDK4, CDKN1A, CDKN2C, and CDKN2D at the 0.05 level. The SNPs tested in this study were selected to tag all the common variants in each gene and not because of their putative functional effects. However, we were unable to design assays for 31 of the selected tSNPs, and 35 of the 199 common SNPs were tagged with $r^2 < 0.8$. Furthermore, although tSNPs based on HapMap and EGP data are likely to tag most of the common SNPs, there is a possibility that other unknown common variants were poorly tagged.

There have been many candidate SNP/gene association studies for ovarian cancer published over the past few years. Few of the published studies report results that are statistically significant; but most of these studies have had insufficient statistical power to detect moderate risks even for common genetic variants. Furthermore, very few studies have used comprehensive tagging approaches to capture all the common variation in a gene. Where associations have been found, the existence of a susceptibility allele remains unproven because results await confirmation, or there are conflicting results in follow-up studies. The current study illustrates the value of large consortia to follow-up putative positive associations for common polymorphisms in complex diseases to
clarify their risks. In the future, a consortium approach to genetic epidemiology studies might also enable the analysis of rarer genetic variants and gene-gene/environment interactions where individual studies have inadequate power.

Acknowledgments

Received 9/1/2006; revised 10/27/2006; accepted 2/1/2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Joan MacIntosh, Hannah Munday, Barbara Perkins, Clare Jordan, Kristy Driver, Nital Shah, the local general practices and nurses and the East Anglian Cancer Registry for registration of the United Kingdom cases; the EPIC-Norfolk investigators for recruitment of the United Kingdom controls; Craig Lucarini and Don Conroy for expert technical assistance. Dr. Stephen Chancok of the Core Genotyping Facility, Division of Cancer Epidemiology and Genetics of the National Cancer Institute (United States), Dr. Neelona Szeszenia-Dabrowska of the Nofer Institute of Occupational Medicine (Lodz, Poland), and Dr. Witold Zatonski of the M. Sklodowska-Curie Institute of Oncology and Cancer Center (Warsaw, Poland) for their contribution to the Polish Breast Cancer Study; Shari Hennesson, Caryllyn Jackson, Jessica Kopic, and Mary Hartley for the HOPE project; Karin Goodman and900

www.aacrjournals.org 3035 Cancer Res 2007; 67: (7). April 1, 2007

SNPs in Cell Cycle Control Genes and Ovarian Cancer Risk

References


Tagging Single Nucleotide Polymorphisms in Cell Cycle Control Genes and Susceptibility to Invasive Epithelial Ovarian Cancer

Simon A. Gayther, Honglin Song, Susan J. Ramus, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/67/7/3027

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2007/03/27/67.7.3027.DC1

Cited articles
This article cites 40 articles, 17 of which you can access for free at:
http://cancerres.aacrjournals.org/content/67/7/3027.full.html#ref-list-1

Citing articles
This article has been cited by 22 HighWire-hosted articles. Access the articles at:
/content/67/7/3027.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.