A Cross-Cancer Genetic Association Analysis of the DNA Repair and DNA Damage Signaling Pathways for Lung, Ovary, Prostate, Breast, and Colorectal Cancer

Peter M. Scarbrough1,2,3, Rachel Palmieri Weber1,2, Edwin S. Iversen4, Yonathan Brhane5, Christopher I. Amos6, Peter Kraft7, RayJean J. Hung5, Thomas A. Sellers8, John S. Witte9,10, Paul Pharoah11,12, Brian E. Henderson13,14, Stephen B. Gruber15,14, David J. Hunter7, Judy E. Garber15, Amit D. Joshi7, Kevin McDonnell13,14, Doug F. Easton16, Ros Eeles17,18, Zsofia Kote-Jarai17,18, Kenneth Muir19,20, Jennifer A. Doherty5, and Joellen M. Schildkraut1,2,21

Abstract

Background: DNA damage is an established mediator of carcinogenesis, although genome-wide association studies (GWAS) have identified few significant loci. This cross-cancer site, pooled analysis was performed to increase the power to detect common variants of DNA repair genes associated with cancer susceptibility.

Methods: We conducted a cross-cancer analysis of 60,297 single nucleotide polymorphisms, at 229 DNA repair gene regions, using data from the NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) Network. Our analysis included data from 32 GWAS and 48,734 controls and 51,537 cases across five cancer sites (breast, colon, lung, ovary, and prostate). Because of the unavailability of individual data, data were analyzed at the aggregate level. Meta-analysis was performed using the Association analysis for SubSets (ASSET) software. To test for genetic associations that might escape individual variant testing due to small effect sizes, pathway analysis of eight DNA repair pathways was performed using hierarchical modeling.

Results: We identified three susceptibility DNA repair genes, RAD51B (P < 5.09 x 10^-6), MSH5 (P < 5.09 x 10^-6), and BRCA2 (P = 5.70 x 10^-4). Hierarchical modeling identified several pleiotropic associations with cancer risk in the base excision repair, nucleotide excision repair, mismatch repair, and homologous recombination pathways.

Conclusions: Only three susceptibility loci were identified, which had all been previously reported. In contrast, hierarchical modeling identified several pleiotropic cancer risk associations in key DNA repair pathways.

Impact: Results suggest that many common variants in DNA repair genes are likely associated with cancer susceptibility through small effect sizes that do not meet stringent significance testing criteria. Cancer Epidemiol Biomarkers Prev; 25(1): 1-8. ©2015 AACR.

Introduction

DNA damage is an established mediator of carcinogenesis (1). Several carcinogens (e.g., chemical mutagens, viruses, and irradiation) are known to cause cancer through their ability to damage DNA (2–6). Consistent with this established model of carcinogenesis, mutations in many genes known to confer cancer risk [e.g., TP53 (7), ATM (8), BRCA1 (9), BRCA2 (10)], are known to play major roles in DNA damage repair and signaling...
response (11–15). However, although mutations in these genes are associated with high degrees of individual cancer risk (7, 9, 10), these rare events explain only a small fraction of all cancers (5). Given the importance of DNA damage to carcinogenesis, it is plausible that cancer risk would be conferred by common variants of these and other DNA repair genes, and that this risk could be measured in large, genome-wide association studies (GWAS).

GWAS have identified hundreds of single nucleotide polymorphisms (SNP) and susceptibility loci associated with risk for various cancers (16–26). However, few GWAS have identified cancer susceptibility loci near DNA repair genes at stringent levels of significance that have also been shown to function through altered DNA repair (21, 24, 26, 27). These data suggest that common variants in DNA repair genes may not make important contributions to cancer susceptibility, and that cancer susceptibility may be mostly conferred by high-risk, rare variants within this class of genes. However, it is possible that underpowered association studies could miss common variants with weak effect sizes. In order to investigate this hypothesis, a comprehensive candidate gene association study of DNA repair genes was performed.

This study analyzes genetic data from 229 DNA repair genes. In order to increase the power to detect common variant effects, a meta-analysis was performed, using the NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) Network database, which includes data from breast, colon, lung, ovary, and prostate cancer. The Association analysis for SubSETs (ASSET) software package (Bioconductor) was used to conduct the meta-analysis of the large dataset (48,734 controls, 51,537 cases), which also allows for the evaluation of subset effects in a potentially heterogeneous dataset. Because the effect for each SNP may only reach significance in certain cancers (a subset of studies), this represents a powerful and practical approach to meta-analysis. The use of a candidate gene study restricted to DNA repair genes, the size and comprehensiveness of the GAME-ON database, and the use of ASSET to interrogate this large dataset for subset effects with minimal loss of power, represents a significantly more powerful approach to detect individual genetic variants in loci near DNA repair genes than has been previously attempted.

In order to test for cancer risk associations among DNA repair genes, which might escape individual variant testing due to weak effect sizes, dimensional reduction of the dataset was also performed by pathway analysis, using hierarchical modeling (28, 29). DNA repair genes segregate into fairly exclusive, well-defined pathway categories, which provides a strong, rational basis to use this information as a means to achieve dimensional reduction of the dataset, as findings in the pathway categories are therefore more likely to have underlying biologic meaning and less likely to be an artifact of pathway analysis procedures. The hierarchical modeling procedure was selected for use in this study because of its compatibility with the summary-level data available in the GAME-ON database and because this approach to pathway analysis uses information from across the entire dataset, instead of being driven by only a handful of the most significant individual variants. Using pathway membership as binary covariates, the multivariate regression framework of hierarchical modeling allowed for estimation of pathway effect size and significance (P-value) for each pathway. Significant effects in the pathway covariates were interpreted as supportive evidence for the associations between variants in the DNA repair pathways and cancer susceptibility.

**Materials and Methods**

**Study population**

The GAME-ON Network (http://epi.grants.cancer.gov/gameon/) includes GWAS data from 32 studies across North America and Europe as well as Australia, representing five common cancer sites: breast, colon, lung, ovary, and prostate (16, 17, 19–23, 30–33). In total, this included 51,537 cancer cases and 48,734 controls. Data analyzed included summary statistics for each study, after adjusting for age, gender, and population stratification using principal components as applicable (Supplementary Table S1). Genomic variant data was imputed to the 1000 Genomes reference panel using either MACH or IMPUTE (34–36). Imputation was separately carried out for each cancer site. Following imputation, there were 6,300,179 SNPs available for analysis, which were shared among all the GAME-ON databases. To avoid population stratification, all study participants included in the analysis were of European descent. Table 1 summarizes the sample sizes of each participating study, and more detailed characteristics are provided in Supplementary Table S1.

**Gene, SNP, and pathway selection**

We initially identified 247 DNA damage repair and signaling response genes using Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database (http://www.genome.jp/kegg/pathway.html). Because the GAME-ON data did not include sex chromosomes, the gene list was reduced to 229 genes for the final analysis (Supplementary Table S2).

Single nucleotide polymorphisms (SNP) were queried from the region of every gene included in the study, using dbSNP (http://www.ncbi.nlm.nih.gov/snp/) and the GRCh38 reference build of the human genome. There is no data to suggest how far a SNP may be from a gene and still have functional effects on that gene. It is known that variants that affect gene activity can be located as far as 100 kb away from the start and stop sites of the genes; however, the inclusion of a larger search window reduces study power and increases the chance that associations are found that apply to genes other than the one of interest. In an attempt to address these competing concerns, gene regions were a priori defined as 50 kb upstream and downstream of the official start and stop sites for each gene (Supplementary Table S2). Selection criteria for a SNP’s inclusion into the study included MAF > 0.01 and being part of the 1000 Genomes database. This resulted in the initial selection of

| Table 1. Summary of GAME-ON GWAS included in the ASSET Meta-Analysisa |
|-----------------------------|----------|----------|----------|
| Cancer site                | Number of GWAS | Cases (N) | Controls (N) |
| CORRECT: colorectal         | 6        | 5,100    | 4,831    |
| DRIVE: breast              | 11       | 15,748   | 18,084   |
| ER-negative                | 8        | 4,939    | 13,128   |
| ELLIPSE: prostate          | 6        | 14,160   | 12,724   |
| Invasive serous            | 3        | 2,556    | 9,123    |
| Invasive mucinous          | 3        | 306      | 9,123    |
| TRICL: lung                | 6        | 12,160   | 16,838   |
| Adenocarcinoma             | 6        | 3,718    | 15,871   |

Abbreviation: N, number.

aAll cancer cases for a particular site were included in meta-analyses, except where subsets were included or used in place of all cancer cases.
156,804 SNPs. SNPs were then omitted from the analysis if they were not present within every dataset in the GAME-ON network, resulting in a final count of 60,297 SNPs to be included in this study (Supplementary Table S2).

**Statistical analysis**

Because of the unavailability of individual data for all five cancer sites, the data were analyzed at the aggregate level only. All datasets were standardized so that the reference allele for each SNP was the same across datasets. Summary statistics for each genetic variant were obtained for each cancer site and for select histologic subtypes of breast [estrogen receptor (ER)-negative], prostate (aggressive), ovarian (serous and mucinous), and lung (adenocarcinoma) cancers (Table 1). Lung adenocarcinoma, serous ovarian cancer, and mucinous ovarian cancer were subtypes chosen for special inclusion into this study, given their prior associations with DNA repair genes (24, 37, 38). ER-negative breast cancer and aggressive prostate cancer subsets were also included to reflect genetic associations that may be linked with more aggressive forms of the disease. Aggressive prostate cancer was defined as disease cases having a Gleason score of 8 or greater (except for the BPC3 and CGEMS studies which included cases with tumor stage C or greater or cases with a Gleason score of 7 or greater, respectively; ref. 30).

All dataset summary statistics included odds ratios (OR) and standard errors (SE) derived from logistic regression analyses. Study-specific results were combined within each cancer site using a fixed effects model. Pooled estimates by cancer site were adjusted for principal components for population structure, and gender, where applicable.

Using the subset-based approach provided for by the ASSET software package (http://www.bioconductor.org; ref. 39), each genetic variant was evaluated for pleiotropic association with cancer risk across multiple cancer sites and histologic subtype. For every genetic variant, effect sizes between studies were combined, by finding the best subset to maximize the test statistic. The final test statistic for each SNP is obtained by maximizing the subset-specific test statistic over all possible subsets, correcting for multiple testing. ASSET calculates the effect size and significance of each SNP across all studies and also returns a list of studies that constitute the “best subset” of studies associated with the SNP under the assumption of a common direction of association (“1-sided ASSET analysis”) or it can allow for the assumption that significant effects may occur in opposite directions for the same genetic variant, between studies (“2-sided ASSET analysis”). In practice, however, genetic loci that are detected as significant across cancer studies have an overwhelming tendency to have the same direction of association. In this report it was assumed that variants of DNA repair genes were not likely to have effects that were opposite in direction across cancer sites.

The correlation between studies was corrected for by tabulating the number of shared cases and controls between studies and generating a covariance matrix when estimating standard errors. This included overlapping controls from the UK ovarian cancer and UK breast cancer GWAS, both of which included controls form the Wellcome Trust Case Control Consortium (WTCCC). Because significance effects for each SNP may only exist in certain cancers (a subset of studies), this represents a powerful and practical approach to meta-analysis. Of the 60,297 SNPs included in the study, 9,806 SNPs were found to not be in high linkage disequilibrium (LD; R² > 0.70). This SNP count was used to set the threshold for a genetic variant to reach statistical significance, using the Bonferroni correction, \[ P = 5.09 \times 10^{-15} \]. The statistical significance for each genetic variant was calculated using the Bonferroni method in ASSET.

**Hierarchical modeling, pathway analysis**

To reduce the correlation structure in the SNP dataset, SNPs were pruned from the analysis if they were found to be in high LD (R² > 0.70), as determined using the online SNP Annotation and Proxy Search (SNAP) tool (Broad Institute, http://www.broadinstitute.org/mpg/snap/). If LD information for a SNP was not available from the SNAP tool, it was pruned from the analysis. This resulted in 9,806 SNPs available for pathway analyses.

SNP pathway membership was determined based on the DNA repair gene it was linked to and that gene’s membership in DNA damage repair and signaling pathways, as indicated by the KEGG Pathway Database (Supplementary Table S3). As a result, it was possible for a SNP to be a member of more than one pathway. The hierarchical modeling method (28, 29) used was performed in R (R Foundation for Statistical Computing, http://www.R-project.org/; Version 3.1.1, 2014). Briefly, hierarchical modeling was performed using the summary level data from the GAME-ON consortium. First-stage estimates of SNP association with each cancer site (OR, SE, and \( P \)-values < 0.05) were generated by adjusting for principal components, as applicable. This information was then entered into a multivariate regression framework, incorporating higher level information about the SNP (i.e., pathway membership as binary covariates) in order to improve the ranking of results. The effect size and association for each DNA repair pathway covariate was calculated for each cancer site. The SEs were estimated based on the folded-normal distribution (40).

**Results**

Figure 1 illustrates the genomic distribution for all SNPs included in the analysis and the corresponding \( P \)-values for association with cancer risk across one or more cancer sites. Manhattan plots for each of the studies included in the meta-analysis were also generated (Supplementary Fig. S1). After correction for multiple comparisons, 29 genomic markers reached statistical significance. Twenty-six of the 29 SNPs were within the RAD51B gene locus (14q24.1). Three of the 29 statistically significant SNPs were within the MSH5 gene locus (6p21.33). A single SNP, near the BRCA2 gene locus (13q13.1), reached borderline significance \( (P = 5.70 \times 10^{-8}) \). This SNP was at the edge of the defined gene locus window and was actually located within the FRY gene. Although FRY has been previously associated with prostate cancer risk, it is not directly involved in DNA damage repair (41). The other 168 SNPs within the BRCA2 gene did not reach significance testing criteria.

The SNPs with the lowest \( P \)-value at each locus (RAD51B, MSH5, BRCA2) were then analyzed for pleiotropic association with cancer risk (Table 2). RAD51B-associated marker, rs11844632, had an overall (pleiotropic) OR of 0.90 [95% confidence interval (CI), 0.88–0.93; \( P = 5.46 \times 10^{-12} \)] across multiple cancer sites. The highly significant inverse association was limited to breast cancer \( (P = 8.14 \times 10^{-5}) \), ER-negative breast cancer \( (P = 0.01) \), overall prostate cancer \( (P = 1.81 \times 10^{-4}) \), aggressive prostate cancer \( (2.46 \times 10^{-3}) \), and colon cancer
Table 2. Summary statistics of top genetic variants at the RAD51B (14q24.1), MSH5 (6p21.33), and BRCA2 (13q13.1) loci

<table>
<thead>
<tr>
<th>rsID</th>
<th>Context</th>
<th>OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1844632</td>
<td>(14q24.1)</td>
<td>1.09 (0.98-1.00)</td>
<td>0.64 x 10^{-2}</td>
</tr>
<tr>
<td>rs315672</td>
<td>(6p21.33)</td>
<td>1.04 (0.97-1.00)</td>
<td>0.52</td>
</tr>
<tr>
<td>rs56404467</td>
<td>(13q13.1)</td>
<td>1.03 (0.95-1.06)</td>
<td>0.82</td>
</tr>
</tbody>
</table>

(P = 0.01). Associations with lung cancer and ovarian cancer were in the opposite direction of effect and not statistically significant. MSH5-associated marker, rs315672, had an overall (pleiotropic) OR of 1.18 (95% CI, 1.12-1.24; P = 2.53 x 10^{-8}). The marker had a highly significant association with lung cancer (P = 3.99 x 10^{-11}) and had weaker associations with colon (P = 0.051), ovarian cancer (serous subtype; P = 0.050), and lung (adenocarcinaoma subtype; P = 0.03) cancer. BRCA2-associated marker, rs56404467, was borderline significant, having an overall (pleiotropic) OR of 1.39 (95% CI, 1.21-1.59; P = 5.70 x 10^{-6}), driven by an association with overall lung cancer (P = 2.14 x 10^{-6}), colon cancer (P = 7.33 x 10^{-6}), and a weaker association with lung adenocarcinoma (P = 0.01).

To examine whether genomic variations in DNA repair genes might have small, but consistent, effects across cancer sites, left undetected due to being sub-genome wide significant, Q-Q plots were generated using the SNP data from the DNA repair gene regions, from each cancer dataset (Fig. 2). Breast, prostate, and lung (overall and the adenocarcinoma subtype) cancer each showed deviations in P-value distribution greater than would be expected by chance, suggesting small but consistent effects in DNA repair genes may exist. Analysis of the genomic inflation factor (λ) was also performed on each cancer site database (42). A standard allelic test for association was performed, based on the median of the χ2 distribution with df = 1. The λ values produced a modest deviation from the expected value of 1, consistent with the Q-Q plots and also suggestive of an excess number of significant associations in some of the cancer sites. The λ values for each dataset are as follows: breast = 1.10, breast (ER-negative) = 0.96, colon = 0.98, lung = 1.02, lung (adenocarcinoma) = 1.04, ovarian = 1.02, ovarian (serous) = 1.09, ovarian (mucinous) = 1.02, prostate = 1.17, and prostate (aggressive) = 1.08.

In order to statistically model the sub-genome wide significant trends between DNA repair pathways and association with cancer risk, dimensional reduction of the GAME-ON dataset was performed via pathway analysis. Site-specific cancer associations with DNA repair pathways were evaluated using hierarchical modeling (Table 3). The analysis included 9,806 SNPs. Analysis of the homologous recombination (HR) DNA repair pathway revealed pleiotropic associations with colon cancer (P = 4.18 x 10^{-4}) and ovarian cancer: overall (P = 1.39 x 10^{-6}), the serous subtype (P = 1.65 x 10^{-6}), and the mucinous subtype (P = 5.00 x 10^{-6}). Mismatch repair (MMR) showed pleiotropic associations with prostate cancer: overall (P = 3.54 x 10^{-4}) and the aggressive subtype (P = 2.76 x 10^{-4}) and lung cancer: overall (4.86 x 10^{-4}) and the adenocarcinoma subtype (P = 8.76 x 10^{-5}). The DNA repair pathway, nucleotide excision repair, also showed a strong association with breast cancer: overall (P = 7.54 x 10^{-5}) and the ER-negative subtype (P = 1.42 x 10^{-3}) and weaker associations with ovarian cancer (P = 8.69 x 10^{-4}), overall lung cancer (P = 0.024), and colon cancer (P = 0.027). All other DNA repair pathways tested showed at least some weaker associations with one or more cancer subtypes (P < 0.05).

Hierarchical modeling's identification of pleiotropic pathway effects in HR and MMR pathways is consistent with the results obtained from individual SNP testing. In particular, RAD51B and BRCA2 are members of the HR pathway and MSH5 is a member of the MMR pathway. In order to determine whether these three loci, or a small number of other highly significant individual loci, significantly influence the overall hierarchical modeling analysis, a sensitivity analysis was performed. In the first sensitivity analysis, the RAD51B, BRCA2, and MSH5 gene data were removed from the dataset and hierarchical modeling was repeated (Supplementary Table S4). In the second sensitivity analysis, any genes containing SNPs that had associations with P < 1 x 10^{-6} were removed from the dataset. This resulted in the removal of 6 genes (RAD51B, MSH5, BRCA2, DCLRE1B, SMEK1, RAD52) from the dataset prior to the hierarchical modeling procedure (Supplementary Table S5). Neither analysis appeared to reveal a significant change to the overall results, suggesting that a small number of highly significant loci were not driving the hierarchical modeling results. This suggests that the hierarchical modeling results were most likely a result of a large number of small effect sizes throughout the dataset.

Discussion

DNA damage and repair are known to be critically important to carcinogenesis and rare mutations in critical DNA repair genes are known to be associated with unusually high cancer risk. However, previous GWAS of common genetic variants (MAF > 0.01) have only identified a handful of statistically significant loci known to function through their effects on DNA repair genes. It was hypothesized that this could be due to the inability of even large studies to detect weak effect sizes. This study tested this hypothesis through use of a large heterogeneous database and a flexible meta-

Downloaded from cebp.aacrjournals.org on June 22, 2017. © 2015 American Association for Cancer Research.
analysis strategy, which represents an unprecedented increase in statistical power to detect associations among common variants of DNA repair genes. This analytical strategy was supplemented with a strategy of dimensional reduction of the dataset, through pathway analysis, to also detect evidence of trends of association between cancer risk and common variants that may escape common variant testing by not meeting the genome-wide significance testing criteria.

Our results indicated that the RAD51B locus was strongly associated with breast cancer and contained a weaker association with prostate cancer, although this did not achieve statistical significance. This locus has been previously associated with breast (43–45), prostate (18, 46), and mucinous ovarian cancer risk (24). Of the associated SNPs at RAD51B, two were previously reported in the literature, rs10483813 and rs17828907 (18, 43–45, 47, 48). No associations were detected for mucinous ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).

From the MSH5 locus, although rs3131379 was previously found to be associated with lung cancer (27, 37, 49, 50), this SNP was not included in our analysis (because it was not present in all GAME-ON databases), and rs3115672 was identified as the most significant SNP at this locus instead. It should be noted that the pairwise LD between rs3131379 and rs3115672 is very high (R² > 0.99). Our study strongly associated this locus with lung cancer, with only weaker, nonsignificant associations detected for colon cancer, lung adenocarcinoma, and mucinous ovarian cancer. This gene has been previously associated with lung cancer (27, 37, 39, 50) and non-Hodgkin’s lymphoma risk (OR = 1.16, P = 0.03; ref. 51). Interestingly, this locus has also been associated with ovarian cancer at this locus, but this may be due to the relatively small number of mucinous ovarian cancer cases included in this analysis (n = 306).
individuals suffering from lupus erythematosus (52–54), who themselves are known to be predisposed to non-Hodgkin’s lymphoma and lung cancer, while having reduced rates of other solid cancers (55).

Our results identified a SNP at genetic locus 13q13.1, near the BRCA2 gene. Although mutations to BRCA2 have been known to be associated with multiple cancer types (10, 56, 57), this SNP has not been previously identified as a common variant related to cancer susceptibility. The SNP showed strong association with lung cancer. The SNP was located within the analytical window of the BRCA2 gene (±50 kb) but was within the FRY gene region, which is not a canonical DNA repair gene. Thus, this finding should be interpreted with more caution, as supportive evidence of the association of common variants of DNA repair with cancer. However, the possibility that this SNP could affect BRCA2 gene function cannot be ruled out. Furthermore, it represents a potentially novel finding that suggests need for further investigation. This SNP, rs56404467, is in a non-coding exon and likely does not affect the activity or function of the BRCA2 protein but may alter the rate of BRCA2 translation. This contrasts to the smaller and nonfunctional BRCA2 protein resulting from a mutation and could explain the different pattern in cancer associations.

A previous analysis of the BRCA2 gene discovered a locus associated with squamous lung cancer, but this locus was not associated with lung adenocarcinoma, in contrast to our own findings (58). However, secondary analysis identified an additional genetic feature which may explain this discrepancy. There was a different, less significant loci, detected within the BRCA2 gene, but this did not meet the criteria for significance testing of $P < 5.09 \times 10^{-6}$ (rs4942486, $P = 0.003$). We found that this less significant loci was not strongly associated with adenocarcinoma but was associated with overall lung cancer, as previously reported (58). Despite being within the same analytical window, the FRY and BRCA2 loci were over 100,000 bases apart, located within different genes, and did not appear to be in high LD. Therefore, our results support the existence of two separate genetic association loci around the BRCA2 gene.

Overall, individual variant testing failed to find robust evidence for an association between common variants in DNA repair genes and cancer susceptibility. Few loci were identified and all genes had been previously associated with cancer susceptibility. Furthermore, evidence for pleiotropy among common variants in these genetic regions did not receive strong statistical support. However, analysis of Q-Q plots from specific cancer sites, using SNPs data from DNA repair gene regions, suggested that consistent association for common variants in DNA repair genes may exist but are likely difficult to detect due to their small effect sizes. In order to examine this possibility, pathway analysis was used as a tool to reduce the dimensionality of the dataset.

Hierarchical modeling provided statistical evidence that common variants of DNA repair genes are likely associated with cancer susceptibility. Homologous recombination, mismatch repair, and nucleotide excision repair showed strong statistical associations with cancer susceptibility, and for homologous recombination and mismatch repair, this association was present across multiple cancer sites. Sensitivity analysis suggested that these results were not due to the contribution of a few, highly significant loci, but through the combination of small, individual SNP effects throughout the entire dataset.

Figure 2. Observed versus expected $P$-values of DNA repair gene SNPs, by cancer site and overall meta-analysis. SNPs plotted were filtered using the SNAP online tool (see Methods and Materials) and were eliminated from the analysis if $R^2 > 0.70$. Black dots indicate $P$ values from indicated dataset; green lines indicate 95% pointwise expected interval under the null hypothesis; and red line indicates expected observations under the null hypothesis.
A limitation of our analyses is due to the availability of only aggregate summary-level data. Thus, we were unable to evaluate associations with nonaggressive prostate or ER-positive breast cancers. Lack of individual level data also made it difficult to enforce a consistent definition of aggressive prostate cancer. Despite this, our findings support further exploration of associations with DNA repair genes in these subgroups.

The results from pathway analysis and individual loci testing clarify the scientific model of the association of common variants in DNA repair genes with cancer risk. Although rare variants in these genes are known to be strongly linked to cancer incidence, very few individual loci were detected in our analysis, even when using a large database and a powerful analytical approach. Robust statistical significance was only detected under pathway analysis, and was observed to be likely due to the contribution of small effect sizes from multiple genes in DNA repair pathway. These data suggest that common variants of DNA repair genes are associated with cancer risk, but that the associations tend to be weak. These results and their interpretation seem particularly plausible, given the epidemiological observation that mutations at some DNA repair genes have profound deleterious effects (Fanconi anemia, xeroderma pigmentosum, ataxia telangiectasia, etc.). Thus, there is a strong theoretical justification for why common variant effects on cancer predisposition in these genes may be difficult to detect, as they likely face strong, negative selection pressure. This observation provides further rationale for conducting future targeted sequencing to explore the role that rare variants play in determining cancer risk.

Disclosure of Potential Conflicts of Interest

J.E. Garber reports receiving a commercial research grant from Myriad Genetic Labs and Novartis, and is a consultant/advisory board member for Pfizer, SV Life Sciences, and Sequenom. R.A. Eeles has received speakers bureau honoraria from Succinct Communications and has provided expert testimony for Janssen. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions


Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P.M. Scarbrough, R.P. Weber, E.S. Iversen, Y. Bhiane, C.I. Amos, P. Kraft, R.J. Hung, P. Pharoah, D.I. Hunter, D.F. Easton, J.M. Schildkraut


Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P.M. Scarbrough, R.P. Weber, C.I. Amos, P. Kraft, S.B. Gruber, A.D. Joshi, K. McDonnell, R. Eeles

Study supervision: R.J. Hung

Acknowledgments

The authors thank Dr. Nilanjan Chatterjee (NHRI) for his helpful advice and comments on their implementation of the ASSET software.

Grant Support

The scientific development and funding for this project were supported by the following: the Genetic Associations and Mechanisms in Oncology (GAME-ON); an NCI Cancer Post-GWAS Initiative, U19CA148112 (T.A. Sellers, J.M. Schildkraut, P. Pharoah), U19CA148127 (C.I. Amos), U19CA148107 (S.B. Gruber), U19CA148065 (D.J. Hunter, P. Kraft, D.F. Easton), U19CA148537 (B.E. Henderson), National Cancer Institute grants R01CA176016 (J.M. Schildkraut), R01CA088164 (J.S. Witte), R25CA126938 (J.M. Schildkraut), P30 CA023108 (C.I. Amos), U01CA172298 (J.S. Witte), National Institute of General Medical Science grant P20GM103534 (C.I. Amos), Cancer Research UK grants C490/A16561 (P. Pharoah), C490/A10124 (P. Pharoah), C490/A10119 (P. Pharoah), and C1287/A16563 (D.F. Easton).

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.

Received June 16, 2015; revised September 18, 2015; accepted October 5, 2015; published OnlineFirst December 4, 2015.

References


Cancer Epidemiology, Biomarkers & Prevention

A Cross-Cancer Genetic Association Analysis of the DNA Repair and DNA Damage Signaling Pathways for Lung, Ovary, Prostate, Breast, and Colorectal Cancer


Cancer Epidemiol Biomarkers Prev  Published OnlineFirst December 4, 2015.

Updated version  Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-15-0649

Supplementary Material  Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2015/12/04/1055-9965.EPI-15-0649.DC1

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.