UCSF
UC San Francisco Previously Published Works

Title

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

Permalink

https://escholarship.org/uc/item/6p23q2nt

Journal

Cancer Epidemiology Biomarkers & Prevention, 25(1)

ISSN

1055-9965

Authors

Scarbrough, Peter M
Weber, Rachel Palmieri
Iversen, Edwin S
et al.

Publication Date

2016

DOI

10.1158/1055-9965.epi-15-0649

Peer reviewed

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. Scarbrough[1,2,3], Rachel Palmieri Weber[1,2], Edwin S. Iversen[4], Yonathan Brhane[5], Christopher I. Amos[6], Peter Kraft[7], Rayjean J. Hung[5], Thomas A. Sellers[8], John S. Witte[9,10], Paul Pharoah[11,12], Brian E. Henderson[13,14,**], Stephen B. Gruber[13,14], David J. Hunter[7], Judy E. Garber[15], Amit D. Joshi[7], Kevin McDonnell[13,14], Doug F. Easton[16], Ros Eeles[17,18], Zsofia Kote-Jarai[17,18], Kenneth Muir[19,20], Jennifer A. Doherty[6], and Joellen M. Schildkraut[1,2,21,*]

[1]Department of Community and Family Medicine, Duke University Medical Center, Durham, 27710, NC, USA

[2]Cancer Prevention, Detection, and Control Research Program, Duke Cancer Institute, Durham, 27710, NC, USA

[3]Department of Pharmacology and Cancer Biology, Duke University School of Medicine, Durham, NC, 27710, USA

[4]Department of Statistical Science, Duke University, Durham, NC, 27708, USA

[5]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, M5G 1X5, Canada

[6]Geisel School of Medicine, Dartmouth College, Lebanon, NH, 03766, USA

[7]Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, MA, 02115, USA

[8]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, 33612, USA

[9]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, 94143, USA

[10]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, 94158, USA

[11]Department of Oncology, University of Cambridge, Cambridge, CB2 0RE, United Kingdom

[*]Corresponding Author: Joellen Schildkraut, Current Address: University of Virginia, Department of Public Health Sciences, Charlottesville, VA 22908, Mailing address: PO Box 800717, Telephone: (434) 924-8569, Fax: (434) 924-8437, jms2yf@virginia.edu.
[**]Dr. Henderson is recently deceased (June 2015).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[12]Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB2 0RE, United Kingdom

[13]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90089, USA

[14]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, 90089, USA

[15]Cancer Risk and Prevention Clinic, Dana-Farber Cancer Institute, Boston, MA, 02115, USA

[16]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK

[17]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK

[18]Royal Marsden NHS Foundation Trust, London and Sutton, SW3 6JJ, UK

[19]Institute of Population Health, University of Manchester, Manchester, M13 9PL, UK

[20]Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

[21]Department of Public Health Sciences, University of Virginia, Charlottesville, 22908, VA, USA

Abstract

Background—DNA damage is an established mediator of carcinogenesis, though 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 SNPs, 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* ($p < 5.09 \times 10^{-6}$), *MSH5* ($p < 5.09 \times 10^{-6}$) and *BRCA2* ($p = 5.70 \times 10^{-6}$). 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.

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, while 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 (SNPs) 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.

The present 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. Since 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 biological 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 1). 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 1.

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). Since the GAME-ON data did not include sex chromosomes, the gene list was reduced to 229 genes for the final analysis (Supplementary Table 2).

Single nucleotide polymorphisms (SNPs) 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 2). 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 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 2).

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) (30).

All dataset summary statistics included odds ratios (ORs) and standard errors (SEs) 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 age, 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) (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). Since 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) ($R^2 > 0.70$). This SNP count was used to set the threshold for a genetic variant to reach statistical significance, using the Bonferroni correction, $p = 5.09 \times 10^{-6}$. 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 ($R^2 > 0.70$), 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 (Table 3).

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 3). 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 (Supplemental Figure 1). 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 ($p = 5.70 \times 10^{-6}$). This SNP was at the edge of the defined gene locus window and was actually located within the *FRY* gene. While *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% 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 ($p = 8.14 \times 10^{-9}$), ER-negative breast cancer ($p = 0.01$), overall prostate cancer ($p = 1.81 \times 10^{-4}$), aggressive prostate cancer (2.46×10^{-3}), and colon cancer ($p = 0.01$). Associations with lung cancer and ovarian cancer were in the opposite direction of effect

and not statistically significant. *MSH5*-associated marker, rs3115672, had an overall (pleiotropic) OR of 1.18 (95% CI: 1.12–1.24; p = 2.53×10^{-8}). The marker had a highly significant association with lung cancer (p = 3.99×10^{-11}), and had weaker associations with colon (p = 0.051), ovarian cancer (serous subtype) (p = 0.050), and lung (adenocarcinoma subtype) (p = 0.03) cancer. *BRCA2*-associated marker, rs56404467, was borderline significant, having an overall (pleiotropic) OR of 1.39 (95% CI: 1.21–1.61; p = 5.70×10^{-6}), driven by an association with overall lung cancer (p = 2.14×10^{-7}), colon cancer (p = 7.33×10^{-3}), 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, for each cancer dataset (Figure 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 χ^2 distribution with d.f. = 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 = 4.18×10^{-4}) and ovarian cancer: overall (p = 1.39×10^{-6}), the serous subtype (p = 1.65×10^{-6}), and the mucinous subtype (p = 5.00×10^{-5}). Mismatch repair (MMR) showed pleiotropic associations with prostate cancer: overall (p = 3.54×10^{-5}) and the aggressive sub-type (p = 2.76×10^{-3}) and lung cancer: overall (4.86×10^{-4}) and the adenocarcinoma subtype (p = 8.76×10^{-5}). The DNA repair pathway, nucleotide excision repair, also showed a strong association with breast cancer: overall (p = 7.54×10^{-5}) and the ER-negative subtype (p = 1.42×10^{-3}) and weaker associations with ovarian cancer (p = 8.69×10^{-3}), 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 4). In the second sensitivity analysis, any genes containing SNPs that had associations with $p < 1 \times 10^{-4}$, 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 5). 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-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 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 ($R^2 > 0.99$). Our study strongly associated this locus with lung cancer, with only weaker, non-significant 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) (51). Interestingly, this locus has also been associated with individuals suffering from lupus erythematosus (52–54), who themselves are known to be predisposed to non-Hodgkin's lymphoma and lung cancer, while have reduced rates of other solid cancers (55).

Our results identified a SNP at genetic locus 13q13.1, near the *BRCA2* gene. While 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 non-functional 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.

A limitation of our analyses is due to the availability of only aggregate summary-level data. Thus, we were unable to evaluate associations with non-aggressive 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 pigmentosa, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Farmer, PB.; Walker, JM. The Molecular basis of cancer. New York: Wiley; 1985.

2. DeLeo AB, Jay G, Appella E, Dubois GC, Law LW, Old LJ. Detection of a transformation-related antigen in chemically induced sarcomas and other transformed cells of the mouse. Proc Natl Acad Sci U S A. 1979; 76:2420–4. [PubMed: 221923]

3. Kodama K, Ozasa K, Katayama H, Shore RE, Okubo T. Radiation effects on cancer risks in the Life Span Study cohort. Radiat Prot Dosimetry. 2012; 151:674–6. [PubMed: 22908358]

4. Yamagiwa K, Ichikawa K. Experimental study of the pathogenesis of carcinoma. CA Cancer J Clin. 1977; 27:174–81. [PubMed: 406018]

5. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013; 45:1113–20. [PubMed: 24071849]

6. Linzer DI, Levine AJ. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell. 1979; 17:43–52. [PubMed: 222475]

7. Malkin D, Li FP, Strong LC, Fraumeni JF Jr, Nelson CE, Kim DH, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science. 1990; 250:1233–8. [PubMed: 1978757]

8. Swift M, Reitnauer PJ, Morrell D, Chase CL. Breast and other cancers in families with ataxia-telangiectasia. N Engl J Med. 1987; 316:1289–94. [PubMed: 3574400]

9. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990; 250:1684–9. [PubMed: 2270482]

10. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science. 1994; 265:2088–90. [PubMed: 8091231]

11. Kastan MB, Onyekwere O, Sidransky D, Vogelstein B, Craig RW. Participation of p53 protein in the cellular response to DNA damage. Cancer Res. 1991; 51:6304–11. [PubMed: 1933891]

12. Kitagawa R, Kastan MB. The ATM-dependent DNA damage signaling pathway. Cold Spring Harb Symp Quant Biol. 2005; 70:99–109. [PubMed: 16869743]

13. Meyn MS. High spontaneous intrachromosomal recombination rates in ataxia-telangiectasia. Science. 1993; 260:1327–30. [PubMed: 8493577]

14. Moynahan ME, Chiu JW, Koller BH, Jasin M. Brca1 controls homology-directed DNA repair. Mol Cell. 1999; 4:511–8. [PubMed: 10549283]

15. Moynahan ME, Pierce AJ, Jasin M. BRCA2 is required for homology-directed repair of chromosomal breaks. Mol Cell. 2001; 7:263–72. [PubMed: 11239455]

16. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013; 45:392–8. 398e1-2. [PubMed: 23535733]

17. Goode EL, Chenevix-Trench G, Song H, Ramus SJ, Notaridou M, Lawrenson K, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nat Genet. 2010; 42:874–9. [PubMed: 20852632]

18. Joshi AD, Lindstrom S, Husing A, Barrdahl M, VanderWeele TJ, Campa D, et al. Additive interactions between susceptibility single-nucleotide polymorphisms identified in genome-wide association studies and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. Am J Epidemiol. 2014; 180:1018–27. [PubMed: 25255808]

19. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353–61. 361e1-2. [PubMed: 23535729]

20. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012; 131:217–34. [PubMed: 21761138]

21. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144:799–807. e24. [PubMed: 23266556]

22. Pharoah PD, Tsai YY, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet. 2013; 45:362–70. 370e1-2. [PubMed: 23535730]

23. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Hum Mol Genet. 2012; 21:5373–84. [PubMed: 22976474]

24. Earp MA, Kelemen LE, Magliocco AM, Swenerton KD, Chenevix-Trench G, et al. Australian Cancer Study. Genome-wide association study of subtype-specific epithelial ovarian cancer risk alleles using pooled DNA. Hum Genet. 2014; 133:481–97. [PubMed: 24190013]

25. Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis. 2014; 35:1012–9. [PubMed: 24325915]

26. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41:579–84. [PubMed: 19330030]

27. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008; 40:1407–9. [PubMed: 18978787]

28. Brenner DR, Brennan P, Boffetta P, Amos CI, Spitz MR, Chen C, et al. Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls. Hum Genet. 2013; 132:579–89. [PubMed: 23370545]

29. Chen GK, Witte JS. Enriching the analysis of genomewide association studies with hierarchical modeling. Am J Hum Genet. 2007; 81:397–404. [PubMed: 17668389]

30. Amin Al Olama A, Kote-Jarai Z, Schumacher FR, Wiklund F, Berndt SI, Benlloch S, et al. A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease. Hum Mol Genet. 2013; 22:408–15. [PubMed: 23065704]

31. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet. 2009; 41:996–1000. [PubMed: 19648919]

32. Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeboller H, et al. Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. Hum Mol Genet. 2012; 21:4980–95. [PubMed: 22899653]

33. Wang H, Burnett T, Kono S, Haiman CA, Iwasaki M, Wilkens LR, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. Nat Commun. 2014; 5:4613. [PubMed: 25105248]

34. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012; 44:955–9. [PubMed: 22820512]

35. Howie B, Marchini J, Stephens M. (2011) Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1:457–70. [PubMed: 22384356]

36. Marchini J, Howie B, Myers S, McVean G, Donnelly P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet. 2007; 39:906–13. [PubMed: 17572673]

37. Kazma R, Babron MC, Gaborieau V, Génin E, Brennan P, Hung RJ, et al. Lung cancer and DNA repair genes: multilevel association analysis from the International Lung Cancer Consortium. Carcinogenesis. 2012; 33:1059–64. [PubMed: 22382497]

38. Schildkraut JM, Iversen ES, Wilson MA, Clyde MA, Moorman PG, Palmieri RT, et al. Association between DNA damage response and repair genes and risk of invasive serous ovarian cancer. PLoS One. 2010; 5:e10061. [PubMed: 20386703]

39. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, et al. A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am J Hum Genet. 2012; 90:821–35. [PubMed: 22560090]

40. Elandt RC. The folded normal distribution: two methods of estimating parameters from moments. Technometrics. 1961; 3:551–62.

41. Nagai T, Ikeda M, Chiba S, Kanno S, Mizuno K. Furry promotes acetylation of microtubules in the mitotic spindle by inhibition of SIRT2 tubulin deacetylase. J Cell Sci. 2013; 126:4369–80. [PubMed: 23886946]

42. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–8. [PubMed: 18852200]

43. Bhatti P, Doody MM, Rajaraman P, Alexander BH, Yeager M, Hutchinson A, et al. Novel breast cancer risk alleles and interaction with ionizing radiation among U.S. radiologic technologists. Radiat Res. 2010; 173:214–24. [PubMed: 20095854]

44. Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Hum Mol Genet. 2011; 20:4693–706. [PubMed: 21852249]

45. Ma H, Li H, Jin G, Dai J, Dong J, Qin Z, et al. Genetic variants at 14q24.1 and breast cancer susceptibility: a fine-mapping study in Chinese women. DNA Cell Biol. 2012; 31:1114–20. [PubMed: 22313133]

46. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013; 45:385–91. 391e1-2. [PubMed: 23535732]

47. Vachon CM, Scott CG, Fasching PA, Hall P, Tamimi RM, Li J, et al. Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2012; 21:1156–66. [PubMed: 22454379]

48. Warren Andersen S, Trentham-Dietz A, Gangnon RE, Hampton JM, Figueroa JD, Skinner HG, et al. Reproductive windows, genetic loci, and breast cancer risk. Ann Epidemiol. 2014; 24:376–82. [PubMed: 24792587]

49. Doherty JA, Sakoda LC, Loomis MM, Barnett MJ, Julianto L, Thornquist MD, et al. DNA repair genotype and lung cancer risk in the beta-carotene and retinol efficacy trial. Int J Mol Epidemiol Genet. 2013; 4:11–34. [PubMed: 23565320]

50. Zhang M, Hu L, Shen H, Dong J, Shu Y, Xu L, et al. Candidate variants at 6p21.33 and 6p22.1 and risk of non-small cell lung cancer in a Chinese population. Int J Mol Epidemiol Genet. 2010; 1:11–8. [PubMed: 21537448]

51. Lim U, Kocarnik JM, Bush WS, Matise TC, Caberto C, Park SL, et al. Pleiotropy of cancer susceptibility variants on the risk of non-Hodgkin lymphoma: the PAGE consortium. PLoS One. 2014; 9:e89791. [PubMed: 24598796]

52. Hughes T, Adler A, Kelly JA, Kaufman KM, Williams AH, Langefeld CD, et al. Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. Arthritis Rheum. 2012; 64:485–92. [PubMed: 21952918]

53. Fernando MM, Freudenberg J, Lee A, Morris DL, Boteva L, Rhodes B, et al. Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G. Ann Rheum Dis. 2012; 71:777–84. [PubMed: 22233601]

54. Sánchez E, Comeau ME, Freedman BI, Kelly JA, Kaufman KM, Langefeld CD, et al. Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. Arthritis Rheum. 2011; 63:3493–501. [PubMed: 21792837]

55. Bernatsky S, Kale M, Ramsey-Goldman R, Gordon C, Clarke AE. Systemic lupus and malignancies. Curr Opin Rheumatol. 2012; 24:177–81. [PubMed: 22227880]

56. Rebbeck TR, Mitra N, Wan F, Sinilnikova OM, Healey S, McGuffog L, et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. JAMA. 2014; 313:1347–61. [PubMed: 25849179]

57. Wagner JE, Tolar J, Levran O, Scholl T, Deffenbaugh A, Satagopan J, et al. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. Blood. 2004; 103:3226–9. [PubMed: 15070707]

58. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet. 2014; 46:736–41. [PubMed: 24880342]
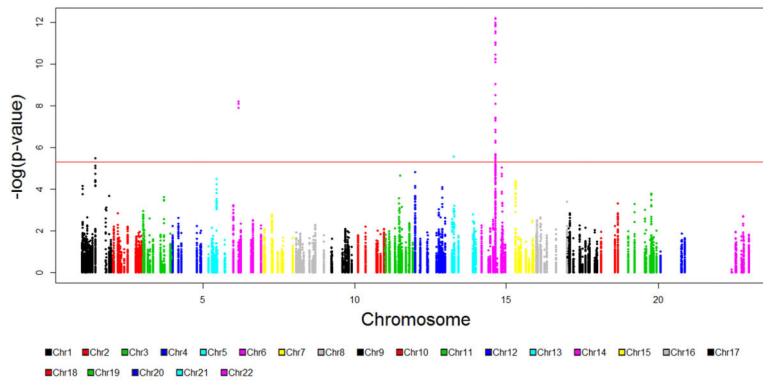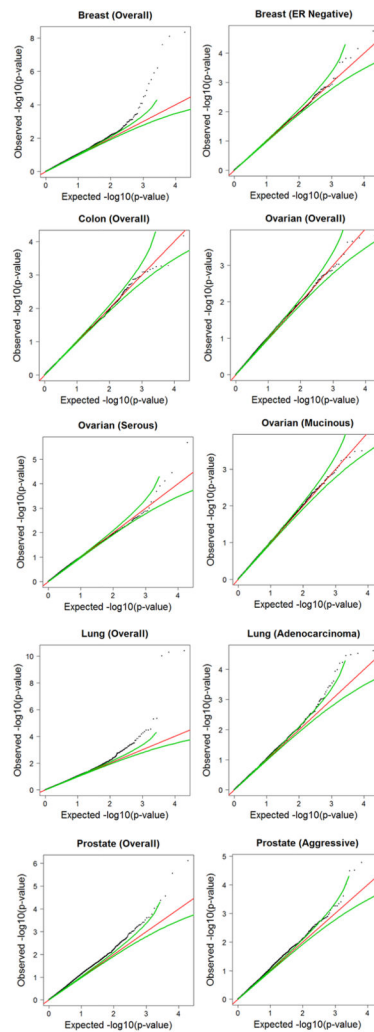
Figure 1.
Manhattan plot, illustrating p-values from 60,297 SNP associations generated from 1-sided ASSET meta-analysis. Statistical significance threshold is denoted by the red line (p = 5.09 $\times 10^{-6}$).

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 = p-values from indicated dataset, green lines = 95% pointwise expected interval under the null hypothesis, red line = expected observations under the null hypothesis.

Table 1

Summary of GAME-ON GWAS included in the ASSET Meta-Analysis[1]

Cancer Site	Number of GWAS	Cases (N)	Controls (N)
CORECT: 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
Aggressive	6	4,450	12,724
FOCI: Ovary	3	4,369	9,123
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

Abbreviations: N, number.

[1] All cancer cases for a particular site were included in meta-analyses, except where subsets were included or used in place of all cancer cases.

Table 2

Summary statistics of top genetic variants at the *RAD51B* (14q24.1), *MSH5* (6p21.33), and BRCA2 (13q13.1) loci.

rs11844632 (14q24.1)	OR (95% CI)	p-value
1-sided ASSET Analysis	**0.90 (0.88–0.93)**	**5.46×10⁻¹²**
Breast (Overall)	**0.90 (0.86–0.93)**	**8.14×10⁻⁹**
Prostate (Overall)	**0.91 (0.87–0.95)**	**2.74×10⁻⁶**
Prostate (Aggressive)	**0.92 (0.86–0.97)**	**2.46×10⁻³**
Bresat (ER Neg)	**0.93 (0.87–0.98)**	**0.01**
Colon (Overall)	**0.92 (0.86–0.98)**	**0.01**
Ovarian (Overall)	1.06 (0.99–1.12)	0.08
Ovarian (Mucinous)	1.04 (0.97–1.12)	0.26
Ovarian (Serous)	1.09 (0.92–1.28)	0.31
Lung (Overall)	0.99 (0.95–1.03)	0.52
Lung (Adenocarcinoma)	1.00 (0.94–1.06)	0.91

rs3115672 (6p21.33)	OR (95% CI)	p-value
1-sided ASSET Analysis	**1.18 (1.12–1.25)**	**2.53×10⁻⁸**
Lung (Overall)	**1.20 (1.14–1.27)**	**3.99×10⁻¹¹**
Lung (Adenocarcinoma)	**1.11 (1.01–1.21)**	**0.03**
Ovarian (Serous)	1.11 (1.00–1.23)	0.05
Colon (Overall)	1.10 (1.00–1.21)	0.05
Ovarian (Overall)	1.06 (0.98–1.15)	0.18
Prostate (Overall)	0.96 (0.91–1.02)	0.21
Breast (ER Neg)	1.05 (0.96–1.15)	0.28
Ovarian (Mucinous)	1.09 (0.86–1.37)	0.48
Prostate (Aggressive)	0.98 (0.89–1.07)	0.59
Breast (Overall)	1.01 (0.95–1.06)	0.82

rs56404467 (13q13.1)	OR (95% CI)	p-value
1-sided ASSET Analysis	**1.39 (1.21–1.61)**	**5.70×10⁻⁶**
Lung (Overall)	**1.43 (1.25–1.64)**	**2.14×10⁻⁷**
Colon (Overall)	**1.31 (1.08–1.60)**	**7.33×10⁻³**
Lung (Adenocarcinoma)	**1.32 (1.06–1.64)**	**0.01**
Breast (Overall)	1.10 (0.95–1.27)	0.19
Breast (ER Neg)	1.16 (0.88–1.54)	0.30
Prostate (Aggressive)	1.15 (0.88–1.50)	0.32
Ovarian (Serous)	1.12 (0.89–1.41)	0.32
Ovarian (Mucinous)	0.85 (0.46–1.56)	0.60
Prostate (Overall)	1.03 (0.87–1.22)	0.71
Ovarian (Overall)	1.03 (0.85–1.24)	0.78

Table 3

Statistical associations (p-values) for cancer risk versus genetic variation within DNA repair pathway.[1]

Cancer (Subtype)	Single Strand Break Repair				Double Strand Break Repair			Other
	Base Excision Repair	Nucleotide Excision Repair	Mismatch Repair	Translesion Synthesis	Fanconi Anemia	Homologous Recombination	Non-homologous end joining	Check Point Factors
Cancer (Subtype)								
Colon (Overall)	**0.016**	**0.027**	0.614	0.342	0.969	**4.18E-04**	0.354	0.546
Breast (Overall)	0.257	**7.98E-05**	0.726	0.431	**0.017**	0.914	0.117	0.496
Breast (ER Negative)	0.587	**1.42E-03**	**0.032**	0.889	0.261	0.542	**3.13E-03**	0.193
Prostate (Overall)	0.401	0.109	**3.57E-05**	0.056	0.232	0.241	0.057	0.217
Prostate (Aggressive)	0.327	0.658	**2.76E-03**	0.455	0.102	0.416	**5.04E-06**	0.555
Ovarian (Overall)	**3.19E-03**	**8.69E-03**	0.552	0.500	0.586	**1.39E-06**	0.276	0.169
Ovarian (Serous)	**0.015**	0.905	0.788	0.772	0.857	**1.65E-06**	0.861	0.769
Ovarian (Mucinous)	0.082	0.076	0.550	0.059	**2.40E-04**	**5.00E-05**	**1.48E-04**	**3.27E-04**
Lung (Overall)	**1.03E-03**	**0.024**	**4.63E-04**	0.885	0.435	0.759	0.181	0.193
Lung (Adenocarcinoma)	0.5313	0.804	**8.50E-05**	**0.020**	0.721	0.268	**2.64E-03**	0.602
Counts by DNA Repair Pathway								
All SNPs	8758	10676	7080	3016	7725	13516	4054	5194
SNAP Filtered SNPs	1568	1850	1174	439	1191	2047	557	880
Genes	37	47	25	13	30	44	15	22

[1]Bolded p-values < 0.05 indicate criteria for significance.