A Field Synopsis on Low-Penetrance Variants in DNA Repair Genes and Cancer Susceptibility

Paolo Vineis, Maurizio Manuguerra, Fotini K. Kavvoura, Simonetta Guarrera, Alessandra Allione, Fabio Rosa, Alessandra Di Gregorio, Silvia Polidoro, Federica Saletta, John P. A. Ioannidis, Giuseppe Matullo

Background Several genes encoding for DNA repair molecules implicated in maintaining genomic integrity have been proposed as cancer-susceptibility genes. Although efforts have been made to create synopses for specific fields that summarize the data from genetic association studies, such an overview is not available for genes involved in DNA repair.

Methods We have created a regularly updated database of studies addressing associations between DNA repair gene variants (excluding highly penetrant mutations) and different types of cancer. Using 1087 datasets and publicly available data from genome-wide association platforms, meta-analyses using dominant and recessive models were performed on 241 associations between individual variants and specific cancer types that had been tested in two or more independent studies. The epidemiological strength of each association was graded with Venice criteria that assess amount of evidence, replication, and protection from bias. All statistical tests were two-sided.

Results Thirty-one nominally statistically significant (ie, $P < .05$ without adjustment for multiple comparisons) associations were recorded for 16 genes in dominant and/or recessive model analyses (*BRCA2*, *CCND1*, *ERCC1*, *ERCC2*, *ERCC4*, *ERCC5*, *MGMT*, *NBN*, *PARP1*, *POLI*, *TP53*, *XPA*, *XRCC1*, *XRCC2*, *XRCC3*, and *XRCC4*). *XRCC1*, *XRCC2*, *TP53*, and *ERCC2* variants were each nominally associated with several types of cancer. Three associations were graded as having "strong" credibility, another four had modest credibility, and 24 had weak credibility based on Venice criteria. Requiring more stringent *P* values to account for multiplicity of comparisons, only the associations of *ERCC2* codon 751 (recessive model) and of *XRCC1* −77 T>C (dominant model) with lung cancer had $P \leq .0001$ and retained $P \leq .001$ even when the first published studies on the respective associations were excluded.

Conclusions We have conducted meta-analyses of 241 associations between variants in DNA repair genes and cancer and have found sparse association signals with strong epidemiological credibility. This synopsis offers a model to survey the current status and gaps in evidence in the field of DNA repair genes and cancer susceptibility, may indicate potential pleiotropic activity of genes and gene pathways, and may offer mechanistic insights in carcinogenesis.

J Natl Cancer Inst 2009;101:24–36

With the rapid increase in the amount of data pertaining to the association of genetic variants with complex diseases comes the challenge to appraise the cumulative evidence. Meeting this challenge is crucial not only to drive research in the field but also to translate results into useful applications for health care and disease prevention (1–4). Although efforts have been made to create synopses for specific fields that summarize all of the data from genetic association studies, including those testing selected variants and those following agnostic genome-wide approaches (5) (http://www.alzforum.org/res/com/gen/alzgene, http://www.schizophreniaforum.org/res/sczgene), such an overview is not available for genes involved in DNA repair.

In the field of DNA repair, the genotypic data that relate to cancer risk have increased exponentially in recent years. This increase derives from an effort to understand how DNA is damaged

Affiliations of authors: Department of Epidemiology and Public Health, Imperial College, London, UK (PV); Institute for Scientific Interchange Foundation, Torino, Italy (MM, SG, AA, FR, ADG, SP, FS, GM); Department of Statistics, Macquarie University, Sydney, Australia (MM); Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece (FKK, JPAI); Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece (JPAI); Center for Genetic Epidemiology and Modeling, Tufts Medical Center, Tufts University School of Medicine, Boston, MA (JPAI); Department of Genetics, Biology and Biochemistry, University of Torino, Italy (GM).

Correspondence to: John P. A. Ioannidis, MD, Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece (e-mail: jioannid@cc.uoi.gr).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djn437

by environmental insults and how the cell's machinery tries to repair the damage without loss of genetic information. Environmental carcinogens such as polycyclic aromatic hydrocarbons, aromatic amines, or *N*-nitroso compounds predominantly form DNA adducts, but they also generate interstrand cross-links and reactive oxygen species, which induce base damage, removal, and single-strand breaks and double-strand breaks (DSBs). Double-strand breaks can also be produced by replication errors or exogenous agents such as ionizing radiation. Unrepaired damage can result in apoptosis (6) or transcriptional changes, and mutations acquired in the process of DNA repair may lead to unregulated cell growth and cancer.

Distinct pathways, each involving numerous factors, have evolved to perform DNA repair (7). The nucleotide excision repair (NER) pathway repairs bulky lesions such as pyrimidine dimers, other products of photochemical reactions, large chemical adducts, and DNA cross-links. The base excision repair (BER) pathway operates on small lesions such as oxidized or reduced bases, fragmented or nonbulky adducts, and adducts produced by methylating agents. At least two pathways for DSB repair exist: homologous recombination and nonhomologous end joining. Mismatch repair (MMR) is an additional category of DNA repair that corrects replication errors (base–base or insertion–deletion mismatches) caused by the DNA polymerase. Finally, alkylated bases are also directly removed by the suicide enzyme methylguanine-DNA methyltransferase.

Genetic variation in some DNA repair genes in each of these pathways appears to influence cancer susceptibility (8,9); however, results pertaining to individual genes have been inconsistent, and an inclusive evaluation of the evidence has not, to our knowledge, been performed. We have collected and regularly updated the cumulative data on associations between polymorphisms in the known DNA repair genes and diverse cancers to create a field synopsis. An online database is maintained at http://www.episat.org, with detailed information on each study included in this synopsis. Here, we present this synopsis and summarize with formal meta-analyses the available data from all studies published before August 31, 2007, that examined associations between a common genetic variant in a DNA repair gene and cancer of any type in humans.

CONTEXT AND CAVEATS

Prior knowledge

Although genetic variation in genes involved in DNA repair may influence susceptibility to cancer and there are many reports of association between individual variants and cancer risk, a comprehensive analysis of genetic association data in this field had not been performed.

Study design

Meta-analysis of reported associations between individual genetic variants and specific cancers using dominant and recessive models of genetic effects.

Contribution

An updateable database and an analytic framework for identifying statistically significant associations and assessing their epidemiological strength in terms of amount of evidence, replication consistency, and protection from bias were developed. The analysis suggested that the vast majority of postulated associations between DNA repair alleles and cancer risk have not been replicated sufficiently to give them strong credibility.

Implications

Possible implications of this work are that larger scale studies would be necessary to establish specific associations of genetic variants in DNA repair and cancer and that the added risk conferred by single variants in DNA repair genes may be small.

Limitations

Biases in genetic association studies could not be fully assessed in this retrospective analysis; the best approach to modeling the genetic effect of a particular variant was not known.

From the Editors

Methods

Literature Search, Selection Criteria, and Data Extraction

We conducted PubMed and HuGE PubLit searches of the English language literature published since 1985. The last update of these searches was in August 2007 when, for purposes of analysis, the databases were frozen. We aimed to identify all published articles in which the frequencies of DNA repair alleles were determined for patients with cancer (of any type) and for unrelated cancer-free control subjects. We excluded highly penetrant mutations (ie, those associated with familial cancer) such as MMR gene mutations in hereditary nonpolyposis colorectal cancer and *BRCA1* and *BRCA2* mutations in familial breast cancer and other familial syndromes. The search terms included all the names or alias of the genes of interest (see Table 1), plus "DNA repair," in combination with terms suggestive of cancer (cancer, neoplasm, tumor, and malignancy). We also excluded data that were unpublished or published in abstracts only. We identified additional articles by searching cited references in the eligible articles. We eliminated obvious overlaps between articles in terms of populations investigated. When publications had overlapping data, we kept the study with the largest sample size. All of the Web tables and references to original articles are available on the DNA repair Web site of the Institute for Scientific Interchange Foundation (http://www.episat.org). The search was performed by M. Manuguerra and independently checked by G. Matullo and F. K. Kavvoura.

From all relevant articles, we collected information on genetic polymorphisms, cancer organ site(s), histological type(s), any exposure evaluated as potential effect modifier (ie, exposures that may interact with genotype), racial descent (Caucasian, Asian, African), and the nature of the recruited cohort (ie, population-based case–control, hospital-based case–control, or case–cohort). Association data were collected as 2 × 2 tables for each polymorphism and cancer type addressed in each study. Two-by-two tables were obtained for all case and control subjects; in addition, separate tables were constructed according to histological type and smoking exposure, whenever such split data were available.

Whenever possible, we used the absolute numbers from published genotype frequencies. When these data were not available, we extracted and used the odds ratios and 95% confidence

intervals as published in the articles. The most common allele was defined as the wild type, unless functional information was available.

Genome-Wide Association Data

Genome-wide association (GWA) studies can test a large number of polymorphisms in an agnostic fashion (ie, without selection for prior credibility). Data from these studies are important to be incorporated in the meta-analyses, when they pertain to relevant DNA repair gene polymorphisms. Therefore, we also searched for GWA data on cancer phenotypes using PubMed, HuGE PubLit, and the National Human Genome Research Institute catalogue of GWA studies. The first search was performed in August 2007; an updated search was performed in July 2008. From the GWA publications identified (10–34), we retrieved all relevant data that were available in the public domain until July 31, 2008. We have thus far been able to retrieve complete datasets for the Cancer Genetic Markers of Susceptibility (CGEMS) study on breast and prostate cancer (http://cgems.cancer.gov/data/). Thus, we selected from the CGEMS database the polymorphisms overlapping with those present in our database (124 and 98 variants for breast and prostate cancer, respectively) and included them in the final database that we used to perform all the meta-analyses. Partial duplication of data with published articles was avoided by excluding those articles from the final meta-analyses. Among the GWA publications for which the datasets were not publicly available, none reported data on specific DNA repair polymorphisms in the main article.

Data Quality Controls

To identify errors in the classification of genotypes, in particular inversion of allele coding, at first screening, we subtracted for every polymorphism the relative genotype frequency of one homozygote from that of the other homozygote, obtaining in the case of inversion a similar difference in frequency but one of opposite sign. To exclude only studies with a high probability of reporting true inversions, we defined a threshold corresponding to a difference of at least 20% between the frequencies of the two homozygous genotypes. We also checked for possible differences among studies due to different genotype frequencies in ethnic groups by looking at dbSNP frequencies reported for the different ethnic groups and at published studies on the same polymorphism when at least three studies were published for the same ethnic group. Twenty-four datasets were excluded from the meta-analyses because they had a very high probability of reporting inversions.

Statistical Analysis

Studies were classified according to type of cancer (ie, the site or organ affected). If at least two different datasets evaluated the same genetic variant and the same type of cancer from at least two different publications, a meta-analysis was performed. The primary analyses combined data on a given association of a genetic variant and a type of cancer. In secondary analyses, separate analyses were performed according to histological type, smoking exposure, racial descent, and the method of subject recruitment.

We used the odds ratio as the metric for all meta-analyses. We explored genotype models based on both recessive and dominant contrasts. If the alleles were A and a, then for a dominant model, a person was classified as 1 if AA and 0 otherwise; for a recessive model with these alleles, a person was classified as 1 if aa and 0 otherwise. If a genetic effect is present, the most appropriate genetic model (recessive, dominant, other) is typically not known for these polymorphisms. Statistically significant results with one model but not with another may occasionally offer a hint to the correct model, but they cannot be taken as proof that the correct genetic model has been identified.

The derived *P* values from these analyses should be interpreted in light of the fact that multiple polymorphisms and two genetic models were analyzed. Therefore, we examined which associations would remain statistically significant if a more stringent threshold, $P \leq .0001$, were adopted. This threshold corresponds to Bonferroni correction for 500 comparisons (the approximate number of associations meta-analyzed [n = 241] multiplied by the number of genetic models [n = 2]). This correction may be too severe, given we performed far fewer meta-analyses for each cancer. Therefore, we also examined which of those associations would attain the threshold of statistical significance after correction for 50 comparisons ($P = .001$) even if the first published study were excluded under the assumption that in genetic epidemiology, the first study often overestimates the effect estimate.

Heterogeneity among the studies was evaluated by Cochran *Q* statistic (35) and was considered statistically significant at *P* less than .10 (36). Both fixed- and random-effects models were used to obtain summary effects. However, because the *Q* test is insensitive in cases where studies are small in size or few in number, we based our main inferences on the random-effects model. This model assumes that the studies are a random sample of a hypothetical population of studies and takes into account within- and between-study variability. We also used the I^2 metric (37) as a measure of the extent of between-study heterogeneity; I^2 values of 50% or higher are considered to reflect large between-study heterogeneity, and values of 25%–50% indicate moderate between-study heterogeneity. With a small number of studies, I^2 can have large uncertainty, so inferences should be interpreted cautiously, and for nominally statistically significant associations, we also estimated the 95% confidence intervals of I^2 (38). We also performed several analyses to explore the possibility for bias. For the formally statistically significant associations, we evaluated whether the results were different after exclusion of the first published study and after adjustment for deviations from Hardy–Weinberg equilibrium (HWE) (39), excluding studies that had statistically significant ($P < .05$) violation of HWE in control subjects according to an exact test. We also evaluated whether smaller studies gave different results than larger studies by using a regression test that formally examined funnel plot asymmetry. The test is a modified version (40) of the original Egger regression test that is considered to correct the inflated type I error of the original regression test. Differences between smaller and larger studies are often interpreted as publication bias, but this is only one possible explanation (41). Such differences may reflect publication bias, other biases, quality differences, or genuine heterogeneity between small and larger studies. We also used the test proposed by Ioannidis and Trikalinos (42) to examine if there was an excess of statistically significant results compared with what one would expect based on

the observed summary effects in each of the meta-analyses. The test was applied to each meta-analysis with nominally statistically significant results and to the whole domain (ie, considering all meta-analyses). We also examined whether there was an excess of studies with statistically significant results in meta-analyses that had found nominally statistically significant summary effects vs those that had nonsignificant summary effects, and in meta-analyses that had large estimated between-study heterogeneity (I^2 > 50%) vs those that did not. The modified regression and excess tests are traditionally considered statistically significant at *P* value less than .10.

Calculations were performed with R, version 2.4.1 (R Foundation for Statistical Computing, Vienna, Austria), and Intercooled STATA, (StataCorp LP, College Station, TX) version 8.2 (College Station, TX). All *P* values are two-sided.

Assessment of Cumulative Evidence

To each nominally statistically significant association, we applied a grading system that was recently developed to assess the strength of the cumulative evidence ["Venice criteria," presented in detail elsewhere (43)]. Briefly, each meta-analyzed association was graded based on the amount of evidence, the extent of replication, and protection from bias. For amount of evidence, a grade of A, B, or C was assigned when the sample size (case and control subjects) for the rarer genotype in the meta-analyses was greater than 1000, 100–1000, or less than 100, respectively. For replication consistency, point estimates of I^2 that were less than 25%, 25%–50%, and greater than 50% were assigned grades of A, B, and C, respectively. For protection from bias, a grade of A means that bias, if present, may change the magnitude but not the presence of an association; a grade of B means that there is no evidence of bias that would invalidate an association, but important information is missing; and a grade of C means that there is a strong possibility of bias that would render the finding of an association invalid. We considered various potential sources of bias, including errors in assigning phenotypes or genotypes, confounding (population stratification), and errors and biases at the level of meta-analysis (publication and other selection biases); errors and biases are also considered in the framework of the observed summary odds ratio estimate. When the summary odds ratio deviated less than 1.15-fold from the null (ie, for odds ratio [OR] values of 0.85–1.15) for meta-analyses based on published data, we concluded that selective reporting bias alone may have rendered the observed association invalid, regardless of whether other biases were present. Therefore, we assigned a grade of C. When the summary odds ratio deviated more than 1.15-fold from the null, a grade of C was given if nominal statistical significance was lost with the exclusion of the first published study or of studies where HWE was violated, or if the results of modified regression or excess tests attained statistical significance, indicating possible bias. In cases where odds ratios deviated more than 1.15-fold from the null, we considered that phenotyping errors could affect the magnitude but not the presence of an effect in this field because the misclassification rate for the various cancers considered here and for the control subjects is unlikely to be that high; genotyping errors were also considered to affect the magnitude but usually not the detection of statistically significant associations in cases where odds ratios exceeded 1.15. Potential confounding from population stratification was considered to have a similar impact (given that at least self-reported racial descent is taken into account in all our analyses). Therefore, a grade of A for protection from bias was assigned if summary odds ratios were greater than 1.15 or less than 0.85, and no bias was detected.

Associations that were assigned three A grades are considered to have strong epidemiological credibility; associations that received a grade of B but for which all other grades were B or greater were considered to have moderate credibility; any association that received a grade of C were considered to have weak credibility.

Results

Main Analyses

Our systematic searches identified 361 articles that referred to cancer risk and DNA repair gene variants (Supplementary Table 1, available online) that examined a total of 1123 associations of gene variants with a type of cancer. Among these, we did not consider for meta-analysis 833 associations where there was only a single dataset available and 50 associations where there were two or more datasets that were all derived from the same article. Ultimately, we performed meta-analyses on 241 associations with a total of 1087 datasets. The summary odds ratio estimates in the dominant and recessive model analyses are shown in Supplementary Tables 2 and 3 (available online). From the 241 analyses, 31 associations involving 16 different genes had a summary effect that was nominally statistically significant, 14 in the dominant model analyses and 17 in the recessive model analyses (Table 1). Four associations were nominally statistically significant in both the recessive and the dominant model analyses. Only 10 of the nominally statistically significant associations involved more than five studies. Of the 31 associations, 19 remained nominally significant after excluding the first published study. Only two of the 31 associations had a *P* value of .0001 or less in the overall analysis: *XRCC1* −77 T>C and lung cancer (dominant model) and *ERCC2* codon 751 and lung cancer (recessive model). Both of these had *P* values slightly below .001 after exclusion of the first published studies (Figure 1).

Secondary Analyses

In general, despite some variability, effect sizes for the associations of a given polymorphism and particular cancers were not statistically significantly different according to histological type and smoking status (Supplementary Tables 4 and 5, available online). Because most studies did not present details and separate data based on these variables, the results should be interpreted with caution. For example, histological information was not available for the association between *XRCC1* −77 T>C and lung cancer (the association with the overall lowest *P* value).

Analyses according to racial descent are shown in Supplementary Table 6 (available online). Of the 31 associations identified in dominant and recessive models, only five were tested in at least two independent studies in at least two different racial descent groups, and the effect sizes for a given association in the different groups did not differ statistically significantly. Moreover, the summary estimates were in the same direction in all racial descent groups, with the exception of the association between *ERCC2*

Table 1. Nominally statistically significant associations of polymorphisms in genes encoding for DNA repair with human cancers at particular sites*

Gene	Polymorphism	Cancer	Model	No. of studies	Sample size†	OR (95% CI)	*P* value	*Q*	*Q* statistic, *P* value	I^2, % (95% CI)	OR (95% CI), excluding first study	*P* value
BRCA2	Codon 1915	Breast	Recessive	2	2566	3.28 (1.78 to 6.06)	.00014	0.74	.39	0	0.50 (0.02 to 14.93)‡	.683
CCND1	Codon 241	Head and neck	Recessive	2	1025	1.82 (1.23 to 2.69)	.003	1.27	.26	21	2.29 (1.34 to 3.93)	.002
ERCC1	Codon 118	Bladder	Dominant	2	1695	0.70 (0.54 to 0.91)	.008	0.01	.94	0	0.71 (0.49 to 1.03)‡	.068
ERCC2	Codon 312	Bladder	Dominant	4	4006	1.20 (1.05 to 1.39)	.009	1.93	.59	0 (0 to 85)	1.16 (0.93 to 1.46)	.189
ERCC2	Codon 312	Lung	Recessive	13	11 469	1.23 (1.06 to 1.43)	.007	13.77	.32	13 (0 to 52)	1.21 (1.02 to 1.43)	.032
ERCC2	Codon 751	Lung	Dominant	18	13 669	1.15 (1.04 to 1.26)	.007	25.30	.09	33 (0 to 62)	1.14 (1.01 to 1.28)	.034
ERCC2	Codon 751	Lung	Recessive	18	13 669	1.26 (1.12 to 1.41)	.0001	11.76	.81	0 (0 to 50)	1.23 (1.09 to 1.39)	.001
ERCC4	Codon 415	Breast	Recessive	6	7685	2.34 (1.17 to 4.69)	.017	3.00	.70	0 (0 to 75)	2.02 (0.97 to 4.22)	.061
ERCC5	Codon 46	Lung	Recessive	2	920	0.60 (0.45 to 0.81)	.001	0.22	.64	0	0.58 (0.40 to 0.82)‡	.002
MGMT	Codon 143	Prostate	Dominant	2	2688	1.22 (1.01 to 1.47)	.042	0.51	.47	0	1.20 (0.99 to 1.46)‡	.058
MGMT	Codon 143	Prostate	Recessive	2	2688	2.02 (1.06 to 3.85)	.033	0.04	.85	0	2.05 (1.06 to 3.98)‡	.030
NBN	Codon 185	Bladder	Dominant	4	4825	1.15 (1.02 to 1.30)	.022	1.08	.78	0 (0 to 85)	1.15 (1.01 to 1.31)	.038
PARP1	IVS9 +104 A>G	Breast	Recessive	2	2467	1.70 (1.06 to 2.71)	.027	0.11	.74	0	1.73 (1.07 to 2.81)‡	.024
POLI	Codon 706	Lung	Dominant	3	3045	1.17 (1.01 to 1.35)	.041	0.67	.72	0 (0 to 90)	1.05 (0.77 to 1.42)‡	.757
TP53	Codon 72	Cervix	Dominant	78	16 575	0.87 (0.78 to 0.98)	.016	150.9	<.01	49 (34 to 61)	0.87 (0.78 to 0.98)	.017
TP53	Codon 72	Lung	Dominant	32	21 477	1.12 (1.03 to 1.23)	.011	60.66	<.01	49 (23 to 66)	1.13 (1.03 to 1.23)	.010
TP53	Codon 72	Lung	Recessive	32	21 477	1.15 (1.01 to 1.30)	.033	50.97	.01	39 (7 to 60)	1.16 (1.02 to 1.32)	.022
TP53	Intron 6 (Msp I)	Breast	Recessive	5	14 030	0.67 (0.51 to 0.88)	.004	4.48	.34	11 (0 to 81)	na	
XPA	23 G>A	Lung	Recessive	8	4032	1.33 (1.12 to 1.57)	.001	6.71	.46	0 (0 to 68)	1.36 (1.13 to 1.64)	.001
XRCC1	−77 T>C	Lung	Dominant	3	3779	1.46 (1.25 to 1.70)	.0000012	1.12	.57	0 (0 to 90)	1.41 (1.16 to 1.72)	.001
XRCC1	Codon 194	Esophageal	Recessive	5	3053	1.46 (1.00 to 2.12)	.048	5.83	.21	31 (0 to 74)	1.65 (1.23 to 2.20)	.001
XRCC1	Codon 194	Head and neck	Recessive	6	2907	2.53 (1.31 to 4.91)	.006	1.91	.86	0 (0 to 75)	2.56 (1.29 to 5.06)	.007
XRCC1	Codon 194	Skin	Dominant	3	662	0.69 (0.50 to 0.96)	.026	2.00	.37	0 (0 to 90)	0.68 (0.42 to 1.10)‡	.114
XRCC1	Codon 194	Stomach	Dominant	4	1539	0.78 (0.62 to 0.98)	.037	1.94	.58	0 (0 to 85)	0.81 (0.61 to 1.07)	.143
XRCC1	Codon 399	Cervix	Recessive	3	3063	1.56 (1.15 to 2.11)	.004	1.96	.37	0 (0 to 90)	na	
XRCC2	Codon 188	Colorectal	Dominant	2	5918	1.16 (1.01 to 1.34)	.034	0.00	1.00	0	1.16 (1.00 to 1.35)‡	.046
XRCC3	4541 A>G (5′ UTR)	Breast	Dominant	4	12 844	1.09 (1.00 to 1.19)	.050	4.08	.25	26 (0 to 72)	1.09 (0.97 to 1.24)	.159
XRCC3	Codon 241	Breast	Recessive	22	32 678	1.09 (1.00 to 1.18)	.039	26.93	.17	22 (0 to 54)	1.08 (0.99 to 1.17)	.092
XRCC3	Codon 241	Stomach	Recessive	5	2153	0.71 (0.52 to 0.97)	.029	2.09	.72	0 (0 to 79)	0.70 (0.51 to 0.96)	.029
XRCC3	IVS7 17893 A>G	Breast	Recessive	4	12 965	0.87 (0.78 to 0.97)	.011	1.59	.66	0 (0 to 85)	0.89 (0.78 to 1.01)	.074
XRCC4	IVS7 −1 A>G	Bladder	Dominant	2	3306	1.27 (1.03 to 1.58)	.026	1.53	.22	35	1.40 (1.13 to 1.74)‡	.002

* The nomenclature of the polymorphisms follows the name used more frequently in the literature: OR = odds ratio; CI = confidence interval; *BRCA2* = breast cancer type 2 susceptibility protein; *CCND1* = cyclin D1; *ERCC*: = excision repair cross-complementing rodent repair deficiency; *MGMT* = O6-methylguanine–DNA methyltransferase; *NBN* = nibrin; *PARP1* = poly (ADP-ribose) polymerase family, member 1; *POLI* = polymerase 1; *TP53* = tumor protein 53; *XPA* = xeroderma pigmentosum, complementation group A; *XRCC* = X-ray repair complementing defective repair in Chinese hamster cells; UTR = untranslated region; na = not applicable (meta-analysis not performed because all studies were published in the same calendar year).

† The sum of cases and controls.

‡ Only one study used to estimate the summary effect.

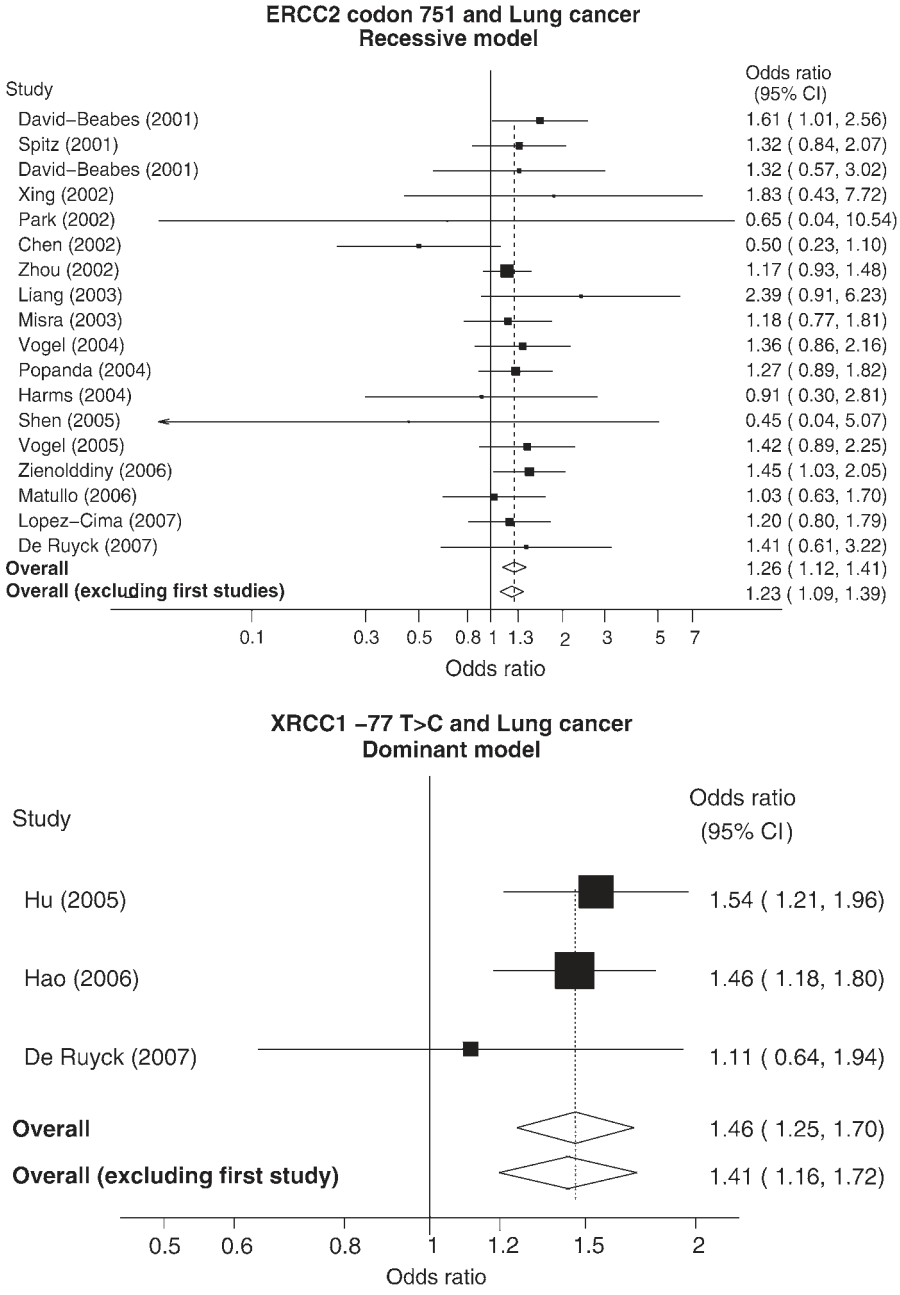


Figure 1. Forest plots for the associations of *ERCC2* codon 751 and lung cancer (recessive model) and *XRCC1* −77 T>C and lung cancer (dominant model). Each study is shown by the odds ratio (**box**) and 95% confidence interval (**horizontal line**). The size of each box is proportional to the weight of each study. Also shown are the **diamonds** of the summary effects based on all studies and excluding the first studies.

codon 751 and lung cancer in the dominant model, where the summary odds ratio was 1.18 (P = .01) in studies of Caucasian populations but was 0.66 (and not statistically significant) in two small studies of subjects of Asian descent. Analyses by racial descent revealed another 12 associations with nominal statistical significance specifically in one racial descent population (Supplementary Table 6, available online).

Heterogeneity

Heterogeneity among studies may be due to gene–environment interactions, gene–gene interactions, study design differences, biases, or chance. Of the 31 associations that were nominally significant in the main analysis, the results of the different studies differed beyond chance (P < .10) for four of them (Table 1), with I^2 estimates suggesting modest amount of heterogeneity.

Across all the 241 meta-analyses using the dominant model, 67 (27.9%) associations had Q test P values that were less than .10. Also, 25 (10.4%) meta-analyses had very large (>75%) estimates of between-study heterogeneity, 46 (19.1%) had large (50%–75%) between-study heterogeneity, and 41 (18.0%) had modest (25%–50%) between-study heterogeneity. In the recessive model, 41 (17.1%) associations had Q test P values less than .10. Eighteen (7.5%) meta-analyses had very large (>75%) estimates of between-study heterogeneity, 26 (10.8%) had large (50%–75%) between-study heterogeneity, and 32 (13.3%) had modest (25%–50%) between-study heterogeneity. Estimates of heterogeneity should

be interpreted cautiously especially when they are based on few studies.

Bias Issues

The 12 associations that were nominally statistically significant in the main analysis but not when the first study was excluded were the following: *BRCA2* codon 1915 and breast cancer with the dominant model, *ERCC2* codon 312 and bladder cancer with the dominant model, *ERCC4* codon 415 and breast cancer with the recessive model, *MGMT* codon 143 and prostate cancer with the dominant model, *POLI* codon 706 and lung cancer with the dominant model, *TP53* intron 6 (Msp I) and breast cancer with the recessive model, *XRCC1* codon 194 and skin and stomach cancer with the dominant model, *XRCC1* codon 399 and cervix cancer with the recessive model, *XRCC3* 4541 A>G (5′ untranslated region [UTR]) and breast cancer with the dominant model, *XRCC3* codon 241 and breast cancer with the recessive model, *XRCC3* IVS7 17893 A>G and breast cancer with the recessive model. Four additional meta-analyses for the dominant model (*ATM* codon 1853 D>N and breast cancer, *TP53* codon 72 and stomach cancer, *XRCC1* codon 399 and leukemia, and *XRCC3* codon 241 and colorectal cancer) and five for the recessive model (*ERCC5* codon 1104 and lung cancer, *MGMT* codon 84 and breast cancer, *TP53* codon 72 and stomach cancer, and *XRCC1* codon 399 and leukemia and prostate cancer) crossed the threshold of nominal significance after exclusion of the first study, but the *P* values were not less than or equal to .001.

After exclusion of studies in which the requirement for HWE was not met, nine of the 31 associations in the main analysis were no longer nominally statistically significant (*BRCA2* codon 1915 and breast cancer with the recessive model, *CCND1* codon 241 and head and neck cancer with the recessive model, *ERCC2* codon 312 and bladder cancer with the dominant model, *ERCC5* codon 46 and lung cancer with the recessive model, *TP53* codon 72 and cervix cancer with the dominant model, *TP53* intron 6 [Msp I] and breast cancer with the recessive model, *XRCC1* codon 194 and esophageal cancer with the recessive model, *XRCC2* codon 188 and colorectal cancer with the dominant model, and *XRCC3* 4541 A>G [5′ UTR] and breast cancer with the dominant model). Conversely, exclusion of HWE-violating studies yielded nominally statistically significant results for three other associations that did not have statistically significant results in the primary analyses (*ERCC2* codon 751 and lymphoma with the dominant model, *TP53* intron 6 (Msp I) and breast cancer with the dominant model, and *ERCC5* codon 1104 and lung cancer with the recessive model; all *P* values were slightly less than .05).

For three of the 31 meta-analyses with nominally statistically significant results in the primary analysis (*TP53* codon 72 and lung cancer in the dominant model analysis, *ERCC2* codon 312 and bladder cancer, and *TP53* codon 72 and cervix cancer in the recessive model analysis), the modified regression test suggested that larger studies had statistically significantly more conservative results than small studies.

For two of these 31 meta-analyses (*TP53* codon 72 with cervical and lung cancer, both in dominant model analysis), there was clear evidence of an excess of individual studies with statistically significant results. Among all the 241 meta-analyses, another 14 (5.8%) had more statistically significant single studies than what would be expected in the dominant model (*CCNH* codon 270 and colorectal cancer; *ERCC2* codon 751 and esophageal and head and neck cancer; *MGMT* codon 84 and head and neck cancer; *TP53* IVS1 −112 G>A and breast cancer; *TP53* codon 72 and breast, cervix, and lung cancer; *XPA* 23 G>A and lung cancer; *XRCC1* codon 194 and head and neck cancer; *XRCC1* codon 399 and breast and colorectal cancer; *XRCC2* codon 188 and breast cancer; and *XRCC3* codon 241 and breast and skin cancer). Six (2.5%) meta-analyses had more statistically significant single studies than what would be expected in the recessive model (*ERCC2* codon 312 and breast cancer, *ERCC2* codon 751 and esophageal cancer, *TP53* IVS1 −112 G>A and breast cancer, *TP53* codon 72 and stomach cancer, *XRCC1* codon 399 and lung cancer, and *XRCC2* codon 188 and breast cancer). These meta-analyses typically pertained to situations where early studies had suggested a statistically significant effect, but an effect in the opposite direction that was also nominally statistically significant was seen (often quite soon) in one or more subsequent studies, reminiscent of the Proteus phenomenon (ie, the rapid interchange of statistically significant results in opposite directions in early published studies) (44).

Among all studies analyzed, we estimated that one would expect an average of 93.8 studies with nominally statistically significant results vs an observed number of 136 ($P = .00002$) for the dominant model analysis; for the recessive model, there would be 85.8 studies expected with nominally statistically significant results vs the observed 100 ($P = .06$). There was an excess of statistically significant results in meta-analyses that had large heterogeneity ($E = 36.2$, $O = 71$, $P = 10^{-6}$, and $E = 21.3$, $O = 42$, $P = 10^{-5}$, in dominant and recessive model analyses, respectively), but not in meta-analyses without large between-study heterogeneity ($E = 57.6$, $O = 65$, $P = .30$, and $E = 64.4$, $O = 58$, $P = .44$, in dominant and recessive model analyses, respectively). According to the dominant model analysis, there was an excess of statistically significant results in meta-analyses with statistically significant results ($E = 17.9$, $O = 31$, $P = .002$) and those with non–statistically significant results ($E = 75.9$, $O = 105$, $P = .001$), whereas no clear excess was seen according to the recessive model analysis for either subgroup.

The majority (61.8%) of the studies analyzed were population-based case–control studies. We did not find a systematic difference in terms of statistical significance between population- and hospital-based studies, although for many associations, data on each type of design were limited or absent (Supplementary Table 7, available online). Population-based studies are considered to be superior in design, but very often the response rate in control subjects is low (50%–60%), with unpredictable implications for the estimates of association. Hospital-based studies pose different problems because response rates are higher, but hospital control subjects may offer a biased representation of the population that gave origin to the case subjects.

In most studies, identification of genetic variants was performed with a 5′ nuclease assay or other recent technologies, and thus, genotyping error should not have caused spurious genetic effects with odds ratios above 1.15 or below 0.85 for common variants (45). Also, the potential for misclassification of phenotypes is low because case–control studies allow accurate disease ascertainment in the field of cancer. Misclassification of control subjects because

of early-stage or undiagnosed cancer was likely to be low, except for the most common cancers and would, if anything, have weakened the observed associations.

Overall Grading and Overview of the Epidemiological Evidence

Based on the Venice criteria, for "amount of evidence," 13 associations were graded as "A," 13 as "B," and five as "C"; for "replication consistency," 24 were graded as "A" and seven as "B"; and for "protection from bias," 10 were graded as "A" and 21 as "C (Table 2)." The main reasons for low protection from bias were the loss of nominal statistical significance after excluding the initial study (n = 12) or violation of the assumption of HWE (n = 9) or the presence of an odds ratio so close to 1 that the nominal association could easily be due to small biases in meta-analyses of published data (n = 3). Overall, three associations (*ERCC2* codon 312 and lung cancer, *ERCC2* codon 751 and lung cancer in recessive model analysis, and *NBN* codon 185 and bladder cancer in dominant model) were assigned a grade of A across all three criteria, and based on these guidelines, they were considered to have strong epidemiological credibility. Another four associations (*ERCC2* codon 751 and lung cancer, *XRCC1* −77 T>C and lung cancer, and *XRCC4* IVS7 −1 A>G and bladder cancer in dominant model, and *XPA* 23 G>A and lung cancer in recessive model analysis) were found to have modest epidemiological credibility, whereas the remaining 24 showed only weak credibility. It is interesting that in analyses limited to populations of Caucasian descent, the association of *ERCC2* codon 751 and lung cancer was also graded as strong. No association was rated as strong in analyses limited to Asian or African populations.

When a more demanding *P* value was required for statistical significance (ie, $P < .0001$), only the *ERCC2* codon 751 association with lung cancer (recessive model) had strong credibility.

Figure 2 presents an overview of the evidence in the field of DNA repair. Because most associations have not been studied with sufficient data, "negative" results should be interpreted cautiously. The evidence seems to be more comprehensive for common cancers where risk is considered to be affected by exposure to environmental carcinogens, such as lung and bladder cancer, and also for breast cancer. Data pertaining to associations are modestly comprehensive for esophageal, head and neck, and colorectal cancer, and less comprehensive for other types of cancer. Some cancers have nominally statistically significant associations with several candidate genes. There are hints that cancers at several sites may

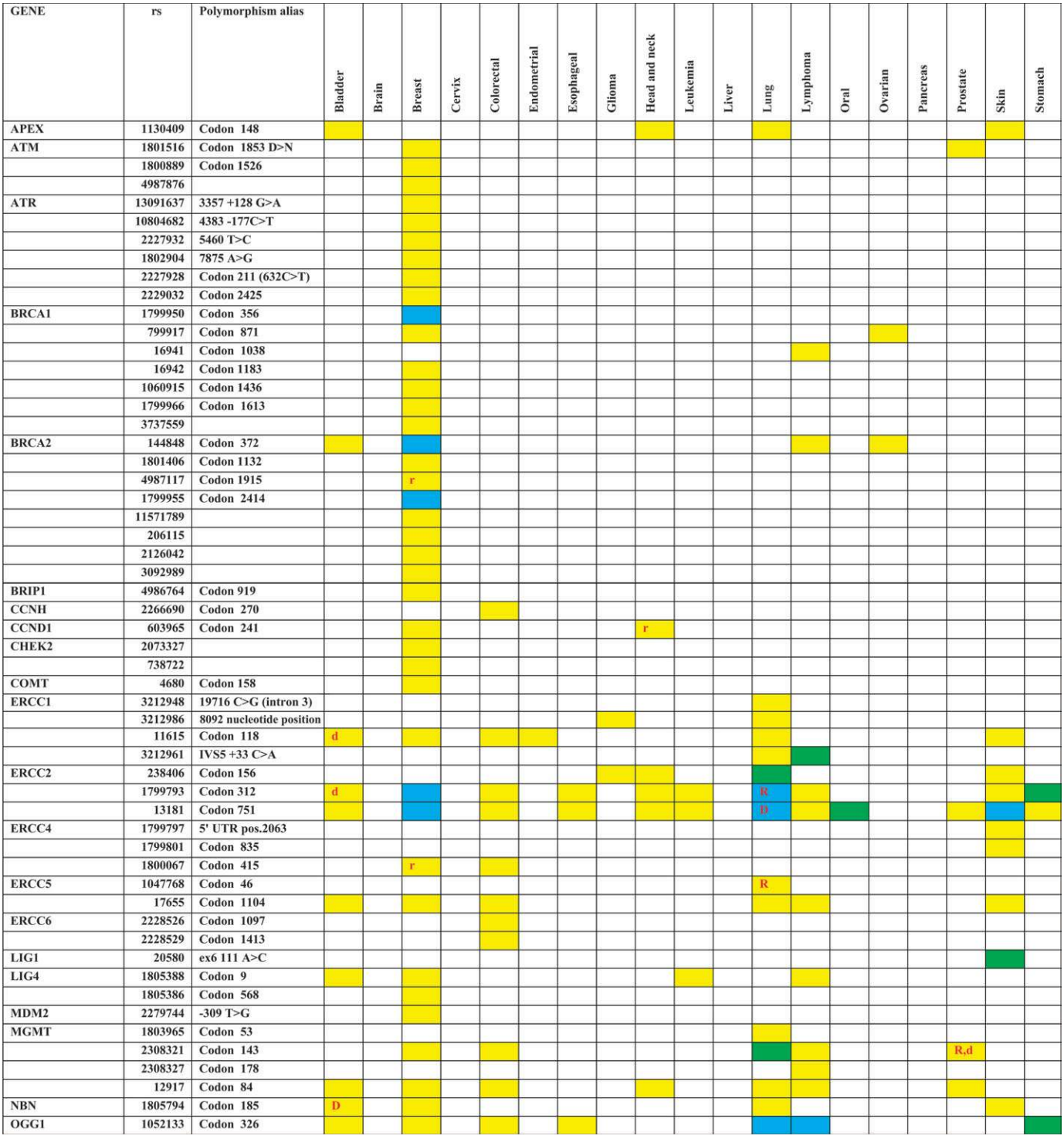
Table 2. Venice grading of the strength of the cumulative epidemiological evidence for the nominally statistically significant associations*

Gene	Polymorphism	Cancer	Model	Protection from bias	Reason	Overall grade
BRCA2	Codon 1915	Breast	Recessive	C	F, HWE	C
CCND1	Codon 241	Head and neck	Recessive	C	HWE	C
ERCC1	Codon 118	Bladder	Dominant	C	F	C
ERCC2	Codon 312	Bladder	Dominant	C	F, HWE, R	C
ERCC2	Codon 312	Lung	Recessive	A		A
ERCC2	Codon 751	Lung	Dominant	A		B†
ERCC2	Codon 751	Lung	Recessive	A		A
ERCC4	Codon 415	Breast	Recessive	C	F	C
ERCC5	Codon 46	Lung	Recessive	C	HWE	C
MGMT	Codon 143	Prostate	Dominant	C	F	C
MGMT	Codon 143	Prostate	Recessive	A		C‡
NBN	Codon 185	Bladder	Dominant	A		A
PARP1	IVS9 +104 A>G	Breast	Recessive	A		C‡
POLI	Codon 706	Lung	Dominant	C	F	C
TP53	Codon 72	Cervix	Dominant	C	HWE, R, E	C
TP53	Codon 72	Lung	Dominant	C	Low OR, E	C
TP53	Codon 72	Lung	Recessive	C	R	C
TP53	Intron 6 (Msp I)	Breast	Recessive	C	HWE	C
XPA	23 G>A	Lung	Recessive	A		B‡
XRCC1	−77 T>C	Lung	Dominant	A		B‡
XRCC1	Codon 194	Esophageal	Recessive	C	HWE	C
XRCC1	Codon 194	Head and neck	Recessive	A		C‡
XRCC1	Codon 194	Skin	Dominant	C	F	C
XRCC1	Codon 194	Stomach	Dominant	C	F	C
XRCC1	Codon 399	Cervix	Recessive	C	F	C
XRCC2	Codon 188	Colorectal	Dominant	C	HWE	C
XRCC3	4541 A>G (5′ UTR)	Breast	Dominant	C	F, HWE, low OR	C
XRCC3	Codon 241	Breast	Recessive	C	Low OR	C
XRCC3	Codon 241	Stomach	Recessive	C	F	C
XRCC3	IVS7 17893 A>G	Breast	Recessive	C	F	C
XRCC4	IVS7 −1 A>G	Bladder	Dominant	A		B†,‡

* Low OR = odds ratio <1.15; R = small-study effect; F = statistical significance lost excluding first study; HWE = statistical significance lost excluding studies violating Hardy–Weinberg equilibrium; E = excess of statistically significant single studies; UTR = untranslated region.

† Did not receive a grade of A for extent of replication.

‡ Did not receive a grade of A for amount of evidence criterion.

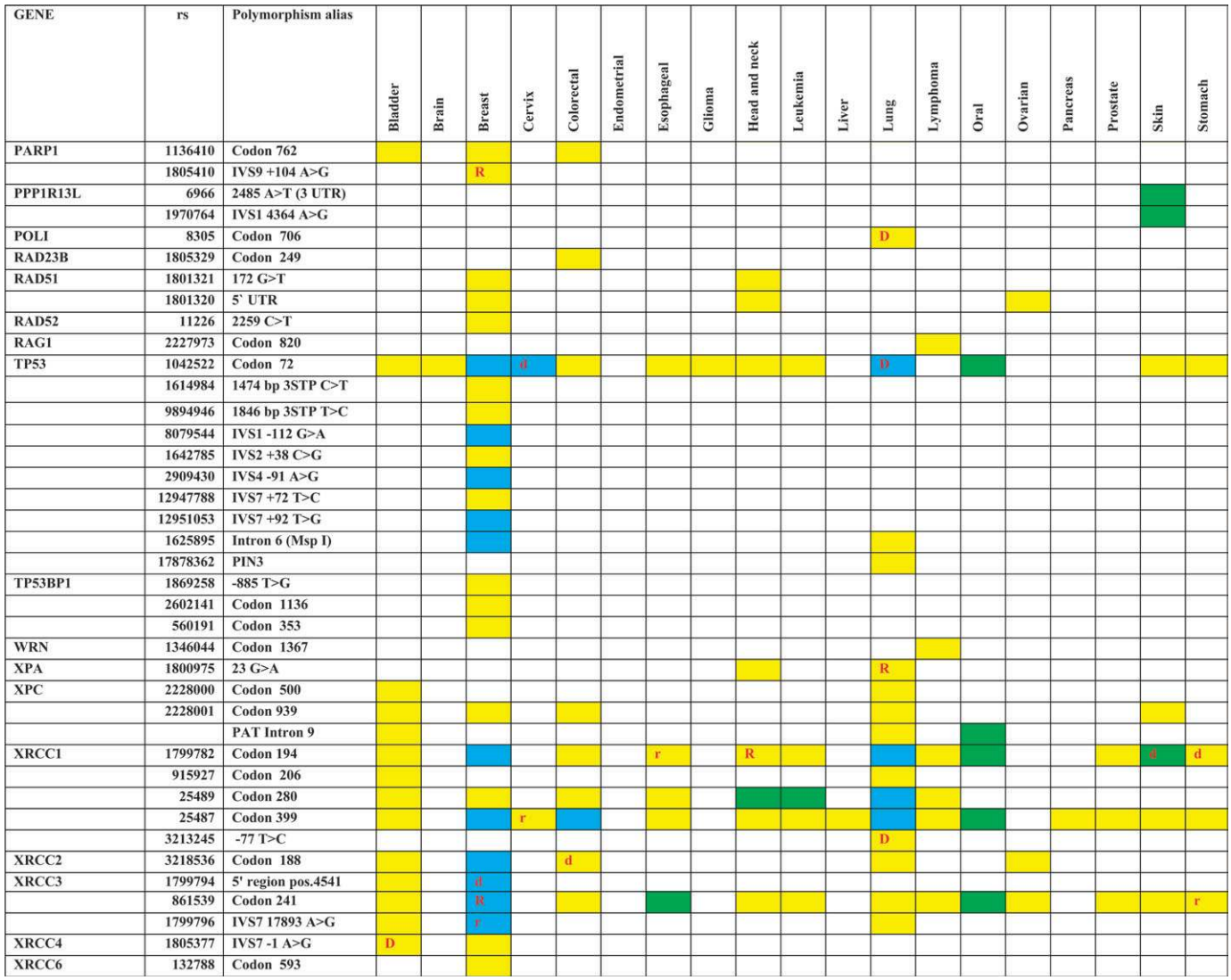


GENE	rs	Polymorphism alias	Bladder	Brain	Breast	Cervix	Colorectal	Endometrial	Esophageal	Glioma	Head and neck	Leukemia	Liver	Lung	Lymphoma	Oral	Ovarian	Pancreas	Prostate	Skin	Stomach
APEX	1130409	Codon 148																			
ATM	1801516	Codon 1853 D>N																			
	1800889	Codon 1526																			
	4987876																				
ATR	13091637	3357 +128 G>A																			
	10804682	4383 -177C>T																			
	2227932	5460 T>C																			
	1802904	7875 A>G																			
	2227928	Codon 211 (632C>T)																			
	2229032	Codon 2425																			
BRCA1	1799950	Codon 356																			
	799917	Codon 871																			
	16941	Codon 1038																			
	16942	Codon 1183																			
	1060915	Codon 1436																			
	1799966	Codon 1613																			
	3737559																				
BRCA2	144848	Codon 372																			
	1801406	Codon 1132																			
	4987117	Codon 1915			r																
	1799955	Codon 2414																			
	11571789																				
	206115																				
	2126042																				
	3092989																				
BRIP1	4986764	Codon 919																			
CCNH	2266690	Codon 270																			
CCND1	603965	Codon 241									r										
CHEK2	2073327																				
	738722																				
COMT	4680	Codon 158																			
ERCC1	3212948	19716 C>G (intron 3)																			
	3212986	8092 nucleotide position																			
	11615	Codon 118	d																		
	3212961	IVS5 +33 C>A																			
ERCC2	238406	Codon 156																			
	1799793	Codon 312	d											R							
	13181	Codon 751												D							
ERCC4	1799797	5' UTR pos.2063																			
	1799801	Codon 835																			
	1800067	Codon 415			r																
ERCC5	1047768	Codon 46												R							
	17655	Codon 1104																			
ERCC6	2228526	Codon 1097																			
	2228529	Codon 1413																			
LIG1	20580	ex6 111 A>C																			
LIG4	1805388	Codon 9																			
	1805386	Codon 568																			
MDM2	2279744	-309 T>G																			
MGMT	1803965	Codon 53																			
	2308321	Codon 143																	R,d		
	2308327	Codon 178																			
	12917	Codon 84																			
NBN	1805794	Codon 185	D																		
OGG1	1052133	Codon 326																			

Figure 2. Overall view of accumulated evidence for association of variants in DNA repair genes and cancer at specific sites. **Colored cells** denote that at least two studies were available and a formal meta-analysis was performed. **Blue color** stands for associations where the total sample size (cases and controls combined) is more than 10 000, **yellow color** stands for associations with 1000–10 000 subjects, and **green color** stands for associations with less than 1000 subjects. For nominally statistically significant associations, the letters **R** and **D** inside the cell denote that the association has nominal statistical significance (*P* value <.05) with recessive and/or dominant model even after exclusion of the first and HWE-deviating studies; the letters **r** and **d** indicate associations that lose their statistical significance in recessive and dominant models, respectively, when the first and/or HWE-deviating studies are excluded. *APEX* = APEX nuclease (multifunctional DNA repair enzyme); *ATM* = ataxia telangiectasia mutated; *ATR* = ataxia telangiectasia and Rad3 related; *BRCA1* = breast cancer type 1 susceptibility protein; *BRCA2* = breast cancer type 2 susceptibility protein; *BRIP1* = BRCA1 interacting protein C-terminal helicase 1; *CCNH* = cyclin H; *CCND1* = cyclin D1; *CHEK2* = CHK2 checkpoint homolog; *COMT* = catechol-*O*-methyltransferase; *ERCC* = excision repair cross-complementing rodent

(continued)

Figure 2 (continued).



repair deficiency; *LIG* = leucine-rich repeats and immunoglobulin-like domains; *MDM2*: = transformed mouse 3T3 cell double minute 2 p53 binding protein homolog (mouse); *MGMT* = O6-methylguanine–DNA methyltransferase; *NBN* = nibrin; *OGG1* = 8-oxoguanine DNA glycosylase; *PARP1* = poly (ADP-ribose) polymerase family, member 1; *POLI* = polymerase 1; *PPP1R13L* = protein phosphatase 1, regulatory (inhibitor) subunit 13 like; *RAD* = RAD homolog B; *RAG1* = recombination activating gene 1; *TP53*bp1 = tumor protein p53 binding protein 1; *TP53* = tumor protein 53; *WRN* = Werner syndrome; *XPA* = xeroderma pigmentosum, complementation group A; *XPC* = xeroderma pigmentosum, complementation group C; *XRCC* = X-ray repair complementing defective repair in Chinese hamster cells; HWE = Hardy–Weinberg equilibrium.

be associated with variants in the same genes, in particular *ERCC2*, *XRCC1*, *XRCC3*, and *TP53*, but most of these associations had weak credibility.

Discussion

This synopsis offers an integrated picture of the accumulated evidence in the field of DNA repair gene variants and cancer risk. The synopsis shows the current status and strength of the available evidence and the gaps in the available data in this field. Only 31 (6%) of the 482 conducted meta-analyses yielded nominally statistically significant results even at a lenient threshold for statistical significance (P = .05), and only 10 of the 31 included more than five datasets. Similar to other areas of genetic associations, many postulated associations were not replicated in the field of DNA repair and cancer. This most likely reflects the presence of a substantial component of false positives and many, perhaps most, of the nominally statistically significant signals that we observed may represent false positives. This may represent a combination of both chance findings and bias, as suggested by some results of testing for excess number of single studies with nominal statistical significance.

The lack of many signals with strong credibility that emerged from our analysis, despite an enormous amount of work in this area over the years, needs careful consideration. The ability of the candidate gene approach to identify genetic risk factors may have been overestimated. Alternatively, the importance of the DNA repair pathway may have been exaggerated. However, there is increasing recognition that genetic risks of cancer conferred by single variants are almost always very modest. This means that

even if the DNA repair pathway is essential for carcinogenesis, extremely large-scale evidence would be necessary to establish with high confidence the presence of specific associations. Environmental and/or lifestyle covariates and genetic interactions may also account for some of the diversity and heterogeneity in the observed results, and capturing this heterogeneity would require studies that carefully collect information for both genetic and environmental variables.

Biological plausibility is difficult to evaluate without clear evidence on the carcinogens involved in the etiology of specific cancers and on the repair pathways that could be plausibly involved. There are a few exceptions, however. The example of *TP53* and lung cancer is particularly intriguing because mutations in *TP53* have been found in lung cancer in association with tobacco smoking (see http://www-p53.iarc.fr/index.html for a systematic database on the subject). Therefore, it is plausible that gene variants for *TP53* could be associated with lung cancer (46,47) if they result in some functional change or if they are in linkage disequilibrium with other functional variants. Another key player in carcinogenesis is the X-ray repair cross-complementing group 1 gene (*XRCC1*), which encodes a scaffold protein within the BER repair system. The lowest *P* value in our field synopsis was obtained for the *XRCC1* −77 T>C polymorphism and lung cancer, although it did not reach an overall strength of grade A after applying the Venice criteria. The *XRCC1* gene has an important role in the BER pathway. A computer analysis predicted that the −77 T>C single nucleotide polymorphism (SNP) was in the core of Sp1-binding motif, which suggested its functional significance (48). Further investigation confirmed that hypothesis and showed that the T>C substitution greatly enhanced the binding affinity of Sp1 to this region, and luciferase assays indicated that the Sp1-high-affinity C-allelic *XRCC1* promoter was associated with a reduced transcriptional activity (48). Other SNPs in *XRCC1* may also be relevant to carcinogenesis (49,50).

When we applied the Venice criteria, three associations (*ERCC2* codon 751 and lung cancer and *ERCC2* codon 312 and lung cancer in recessive model analysis, and *NBN* codon 185 and bladder cancer in the dominant model) were considered to have strong epidemiological credibility, although only the association of *ERCC2* codon 751 and lung cancer also had a *P* value less than or equal to .0001. Contradictory results have been published on the functional implications of these polymorphisms, but computer analyses (PupaSuite: http://pupasuite.bioinfo.cipf.es/) have predicted for all of them an alteration of an exonic splicing enhancer (ESE) sequence. Exonic splicing enhancers appear to be important in exons that normally undergo alternative splicing; different classes of ESE consensus motifs have been described but are not always easily identified. PupaSuite used a script that scans into exon sequences to identify putative ESEs responsive to the human SR proteins SF2/ASF, SC35, SRp40, and SRp55, by using the nucleotide frequency matrices available for them. Moreover, all three of these SNPs with strongly credible associations are located in a region that is conserved between mice and humans.

Some other associations seem to be less strong from an epidemiological perspective, but they provide a focus for future efforts. We have identified several associations that reach less stringent thresholds of statistical significance and we have graded them as having modest or weak credibility. Some of the putative associations that were assigned a grade of C for protection from bias because odds ratios were lower than 1.15 could be real and thus need further investigation. It is increasingly documented that many, possibly most, associations of common variants with complex diseases have very small odds ratios. An odds ratio less than 1.15 has to be seen cautiously in a retrospective meta-analysis, given the unavoidable susceptibility of this design to publication and other reporting biases. However, large-scale prospective investigations may document whether these associations are real.

Several meta-analyses were recently published on DNA repair genes belonging to the DSB repair pathway (51–54) or to the NER pathway, in particular the *ERCC2* gene variants of the latter pathway (54–57). The most recent meta-analysis (55) revealed an increased risk of lung cancer for the *XPD/ERCC2* 751Gln/Gln genotype carriers and a decreased risk for *XPA* 23A carriers. No statistically significant result has been reported for the *XPD/ERCC2* codon 312 polymorphism and lung cancer in either published meta-analysis (54,55), whereas in our updated synopsis, there was a slight but statistically significant increased risk conferred by this allele.

There is considerable evidence that some chemical carcinogens may affect the risk of different types of cancer. For example, alcoholic beverages or food including nitrosocompounds may be involved in head and neck, esophageal, colorectal, and bladder cancer (58). Our checkerboard table approach (Figure 2) may help in understanding if some of these genes are implicated in not only one but in several different types of cancer. The synopsis reveals areas of the DNA repair gene field where sufficient evidence has been accumulated and where it is unlikely that further studies could reveal strong associations. For example, there appears to have been a thorough evaluation of most known gene variants in relation to breast cancer, but all seven nominally statistically significant associations observed were rated as having "weak" credibility. Recent large-scale evaluation in GWA platforms has failed to implicate any DNA repair genes in breast cancer susceptibility, whereas other genes in very different pathways were proposed (11,59). Similar results were obtained in a recent breast cancer pooled analysis (52). Although it is possible that some subtle effects may be missed, even with studies of several thousand subjects, it is likely that the DNA repair gene polymorphisms investigated per se do not play a major role in breast cancer.

Our analyses had some limitations. First, some genuine associations may have been missed due to misclassification from modest nondifferential genotyping or phenotyping error. Second, as in any retrospective meta-analysis of published information, biases can never be fully probed. However, we used an array of diagnostic tests for bias and a consensus approach for grading the evidence, so we believe that our appraisal of the strength of the evidence is not too optimistic. The design of some of the included studies may be problematic or suboptimal in ways that are not possible to see based on the presented information in published reports because reporting in genetic association studies is sometimes deficient in important details (60). This may introduce some heterogeneity

and may create some false-positive signals, but it could also lead to false negatives for some probed associations.

We have created a database that aims to be comprehensive and continuously updated. It is expected that data will continue to accumulate in this field at a rapid pace, and we plan to continue including new studies in our online database and updating our calculations at regular time intervals. In particular, the advent of GWA studies will require the incorporation of their accumulated data in these calculations. Until now, large GWA studies on cancer have been published for breast cancer (11–14), prostate cancer (13,15–20), colorectal cancer (21–25), leukemia (26), lung cancer (27–30), and esophageal cancer (31), melanoma (32,33), and neuroblastoma (34). None of these studies showed highly statistically significant associations for any of these common DNA repair gene variants that would place the DNA repair genes among the few top hits discussed in each of these GWA publications. The genetic effects, if any, are small in magnitude for each implicated polymorphism. Therefore, it is anticipated that even if some of the DNA repair genes are associated with specific cancer types, the signals observed in GWA studies would not necessarily be among the reported low-lying fruit (ie, the polymorphisms with the lowest *P* values). We have so far been able to incorporate data from CGEMS that are publicly available (http://cgems.cancer.gov/), and we will similarly incorporate additional GWA data for other available studies (such as the Genotype and Phenotype database and the Wellcome Trust Case Control Consortium) when the data become publicly available and we have permission to access the data. Such data may help us understand whether DNA repair gene variants affect cancer risk.

Finally, there is some uncertainty as to what would be the best genetic model to represent genetic effects for these variants. We used dominant and recessive models, and the results may differ depending on the model used in nominal statistical significance, especially for associations with weak credibility and borderline *P* values. For functional variants, recessive models may have some rationale because recessive alleles might correspond to the lowest enzymatic activity. Given the nature of the data, we could not examine haplotypes and composite effects involving many genes, whereas data on environmental exposures were typically limited. We recommend that more information on environmental exposures should be routinely collected and reported in these studies. Consortia of investigators performing individual-level analyses extensively covering candidate genes, and considering possible functional variants selected in silico, should also be encouraged.

Despite its limitations, our field synopsis offers a comprehensive picture that would be impossible to obtain from fragmented investigation of single studies or isolated meta-analyses. Building on this evidence base, we can expand, correct, and improve our understanding of the effects of DNA repair genes in the etiology of cancer.

References

1. Lin BK, Clyne M, Walsh M, et al. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol.* 2006;164(1):1–4.
2. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. *Am J Epidemiol.* 1998;148(1):1–3.
3. Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet.* 2006;38(1):3–5.
4. Ioannidis JP, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. *Am J Epidemiol.* 2005;162(4):302–304.
5. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet.* 2007;39(1):17–23.
6. Vispe S, Yung TM, Ritchot J, Serizawa H, Satoh MS. A cellular defense pathway regulating transcription through poly(ADP-ribosyl)ation in response to DNA damage. *Proc Natl Acad Sci USA.* 2000;97(18):9886–9891.
7. Friedberg E, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T. *DNA Repair and Mutagenesis.* Washington, DC: ASM Press; 2006.
8. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst.* 2000;92(11):874–897.
9. Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2002;11(12):1513–1530.
10. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007;39(7):870–874.
11. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447(7148):1087–1093.
12. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci USA.* 2008;105(11):4340–4345.
13. Murabito JM, Rosenberg CL, Finger D, et al. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med Genet.* 2007;8(suppl 1):S6.
14. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet.* 2007;39(7):865–869.
15. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007;39(5):631–637.
16. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007;39(8):977–983.
17. Gudmundsson J, Sulem P, Rafnar T, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet.* 2008;40(3):281–283.
18. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39(5):645–649.
19. Duggan D, Zheng SL, Knowlton M, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst.* 2007;99(24):1836–1844.
20. Eeles RA, Kote-Jarai Z, Giles GG, et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet.* 2008;40(3):316–321.
21. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007;39(8):989–994.
22. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007;39(8):984–988.
23. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet.* 2007;39(11):1315–1317.
24. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 2008;40(5):631–637.
25. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet.* 2008;40(5):623–630.
26. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature.* 2007; 446(7137):758–764.
27. Spinola M, Leoni VP, Galvan A, et al. Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene. *Cancer Lett.* 2007;251(2):311–316.

28. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008;40(5):616–622.
29. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature.* 2008;452(7187):633–637.
30. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008;452(7187):638–642.
31. Hu N, Wang C, Hu Y, et al. Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. *Cancer Res.* 2005;65(7): 2542–2546.
32. Brown KM, Macgregor S, Montgomery GW, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet.* 2008; 40(7):838–840.
33. Gudbjartsson DF, Sulem P, Stacey SN, et al. ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet.* 2008;40(7):886–891.
34. Maris JM, Mosse YP, Bradfield JP, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N Engl J Med.* 2008;358(24): 2585–2593.
35. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med.* 1991;10(11):1665–1677.
36. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med.* 1997;127(9):820–826.
37. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557–560.
38. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ.* 2007;335(7626):914–916.
39. Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP. Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. *Am J Epidemiol.* 2006;163(4):300–309.
40. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med.* 2006;25(20):3443–3457.
41. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ.* 2006;333(7568):597–600.
42. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials.* 2007;4(3):245–253.
43. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol.* 2008 37(1): 120–132.
44. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol.* 2005;58(6): 543–549.
45. Moskvina V, Craddock N, Holmans P, Owen MJ, O'Donovan MC. Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered.* 2006;61(1):55–64.
46. Matakidou A, Eisen T, Houlston RS. TP53 polymorphisms and lung cancer risk: a systematic review and meta-analysis. *Mutagenesis.* 2003; 18(4):377–385.
47. Zhou Y, Li N, Zhuang W, et al. P53 codon 72 polymorphism and gastric cancer: a meta-analysis of the literature. *Int J Cancer.* 2007;121(7): 1481–1486.
48. Hao B, Miao X, Li Y, et al. A novel T-77C polymorphism in DNA repair gene XRCC1 contributes to diminished promoter activity and increased risk of non-small cell lung cancer. *Oncogene.* 2006;25(25):3613–3620.
49. Hu Z, Ma H, Chen F, Wei Q, Shen H. XRCC1 polymorphisms and cancer risk: a meta-analysis of 38 case-control studies. *Cancer Epidemiol Biomarkers Prev.* 2005;14(7):1810–1818.
50. Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am J Epidemiol.* 2005;162(10):925–942.
51. Figueroa JD, Malats N, Rothman N, et al. Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk. *Carcinogenesis.* 2007;28(8):1788–1793.
52. Garcia-Closas M, Egan KM, Newcomb PA, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet.* 2006;119(4):376–388.
53. Han S, Zhang HT, Wang Z, et al. DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies. *Eur J Hum Genet.* 2006;14(10):1136–1144.
54. Manuguerra M, Saletta F, Karagas MR, et al. XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. *Am J Epidemiol.* 2006;164(4):297–302.
55. Kiyohara C, Yoshimasu K. Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: a meta-analysis. *Int J Med Sci.* 2007;4(2):59–71.
56. Benhamou S, Sarasin A. ERCC2/XPD gene polymorphisms and lung cancer: a HuGE review. *Am J Epidemiol.* 2005;161(1):1–14.
57. Hu Z, Wei Q, Wang X, Shen H. DNA repair gene XPD polymorphism and lung cancer risk: a meta-analysis. *Lung Cancer.* 2004;46(1):1–10.
58. de Jong FA, Sparreboom A, Verweij J, Mathijssen RH. Lifestyle habits as a contributor to anti-cancer treatment failure. *Eur J Cancer.* 2008; 44(3):374–382.
59. Breast Cancer Association Consortium. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. *J Natl Cancer Inst.* 2006;98(19): 1382–1396.
60. Yesupriya A, Evangelou E, Kavvoura FK, et al. Reporting of human genome epidemiology (HuGE) association studies: an empirical assessment. *BMC Med Res Methodol.* 2008;8:31.

Funding

This work was made possible by a grant to ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), a network of excellence operating within the European Union 6th Framework Program, Priority 5: "Food Quality and Safety" (contract no. 513943), and by a grant of the compagnia di San Paolo, of the Italian Association for Cancer Research, Italy, and of the Piedmont Region Progetti di Ricerca Sanitaria Finalizzata. F. K. Kavvoura is supported by a PENED training grant cofinanced by EU—European Social Fund (75%) and the Greek Ministry of Development–General Secretariat of Research and Technology (25%).

Notes

The authors declare that they have no competing financial interests. The corresponding author certifies that all authors have agreed to all the contents in the manuscript, including the data as presented. The authors take full responsibility for the study design, data collection, analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

Manuscript received March 27, 2008; revised September 23, 2008; accepted October 30, 2008.