Genotype misclassification in genetic association studies of the rs1042522 *TP53* (Arg72Pro) polymorphism: a systematic review of studies in breast, lung, colorectal, ovarian, and endometrial cancer

Submitted by

Issa J. Dahabreh, MD

In partial fulfillment of the requirements for the degree of

Master of Science

in

Clinical and Translational Science

TUFTS UNIVERSITY

Sackler School of Graduate Biomedical Sciences

May, 2011

ADVISER: Dr. Joseph Lau, MD

Abstract

Background: Genetic association studies in oncology often use tumor tissue as the source of genotyping material for cancer patients. Because tumors exhibit multiple somatic genetic aberrations, tumor-derived genetic material may not accurately represent the individual's constitutional genotype. Loss-of-heterozygosity (LOH) that preferentially affects one variant allele at a locus may affect genotyping results and lead to differential misclassification. We hypothesized that preferential loss of the proline (Pro)-encoding allele at the rs1042522 locus of the *TP53* gene (known as Arg72Pro) may cause such differential misclassification.

Methods: We performed a systematic review of MEDLINE, the Human Genome Epidemiology Literature Finder, the Genetic Association Database and two locus-specific databases dedicated to *TP53* (last search: March 8th, 2011) to identify studies investigating the association of rs1042522 with breast, lung, colorectal, ovarian or endometrial cancer. Data on the populations enrolled, aspects of study design, genotyping methods used and the genotype distribution in cancer cases and controls were extracted from eligible studies. Meta-analysis was performed using both maximum likelihood and Bayesian approaches. For Bayesian analyses, informative priors for the bias effect were derived from a recently published meta-analysis of the same polymorphism in cervical cancer.

Results: We identified 161 case-control studies reported in 134 papers. Sixty-nine studies (28513 cases/31850 controls) investigated breast cancer, 42 (16743 cases/16504 controls) lung cancer, 26 (7377 cases/10011 controls) colorectal cancer, 16 (1982 cases/5226 controls) ovarian cancer and 8 (726 cases/1292 controls) endometrial cancer. Twenty-two

studies (14%) used tumor tissue as the source of genotyping material for cases. Using multilevel models incorporating evidence from all 5 cancers, we found strong evidence that use of tumor tissue was associated with an underestimation of the Pro allele's effect by approximately 20% (OR =0.78; 95% confidence interval, CI, 0.69-0.90). Among studies not using tumor tissue for genotyping we found a significant association between rs1042522 and lung cancer, OR=1.09 (95% CI, 1.02-1.15). In Bayesian analyses, the use of tumor tissue was estimated to lead to an underestimation of the genetic effect by 25%-30%. The probability that use of cancer tissue is associated with underestimation of the Pro allele's "true" effect was more than 95% in all analyses (and higher than 99% in analyses using informative priors). These results were robust to sensitivity analyses.

Conclusions: Use of tumor tissue as the source of genotyping material for cases is associated with significant bias in the estimate of the genetic effect in genetic association studies of rs1042522 with beast, lung, colorectal, ovarian and endometrial cancer. Future studies should avoid the use of tumor tissue as the source for material for genotyping cases, particularly for loci that may be affected by LOH.

Acknowledgements

This work would not have been possible without the help of many individuals. I would like to thank Dr. Thomas Trikalinos for his continued guidance in designing and conducting this project and for his support of my scientific development. I would also like to thank Dr. Christopher Schmid, for his invaluable advice regarding the statistical analyses and interpretation of results, and Dr. Joseph Lau, for his supportive supervision and for providing me with the unique opportunity to work in the highly stimulating environment of the Center for Clinical Evidence Synthesis of the Institute for Clinical Research and Health Policy Studies at Tufts Medical Center. Dr. Vasileia Varvarigou provided indispensable help with many aspects of the systematic review on which the herein presented analyses were based and Dr. Samuel Murray provided useful input for interpreting the biologic significance of the findings. I also owe many thanks to Dr. David Kent for numerous helpful discussions on methodology, both regarding this project as well as other, more general, topics.

Finally, I would like to thank the “Maria P. Lemos” Foundation for providing me with an unrestricted Research Scholarship that partly supported this work.

Table of Contents

List of Tables

List of Figures

Appendices

Appendix Documents

Appendix Document 1: Search strategies.

Appendix Document 2: List of included studies.

Appendix Document 3: Model specification for maximum likelihood and Bayesian analyses.

Appendix Tables

Appendix Table 1: Meta-analysis results for breast cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Table 2: Meta-analysis results for lung cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Table 3: Meta-analysis results for colorectal cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Table 4: Meta-analysis results for ovarian cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Table 5: Meta-analysis results for endometrial cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Table 6: Meta-regression results for breast, lung and colorectal cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Appendix Figures

Appendix Figure 1: Forest plot for studies in breast cancer. Studies have been stratified by the source of DNA for genotyping cases.

Appendix Figure 2: Forest plot for studies in lung cancer. Studies have been stratified by the source of DNA for genotyping cases.

Appendix Figure 3: Forest plot for studies in colorectal cancer. Studies have been stratified by the source of DNA for genotyping cases.

Appendix Figure 4: Forest plot for studies in ovarian cancer. Studies have been stratified by the source of DNA for genotyping cases.

Genotype misclassification in genetic association studies of the rs1042522 *TP53* (Arg72Pro) polymorphism: a systematic review of studies in breast, lung, colorectal, ovarian, and endometrial cancer

Introduction

Lung, breast and colorectal cancer jointly account for the majority of new cancer cases and represent the top three causes of cancer-related mortality in Western countries.[1-2] Breast, ovarian and endometrial cancer are major causes of cancer-related morbidity and mortality among women.[1] Multiple lines of evidence, including studies of familial clustering and studies of heritability in twins, suggest that these common epithelial cancers have a substantial hereditary component.[3] One of the most promising cancer genes for cancer causation is *TP53*, a gene encoding for a 53kD transcription factor involved in regulating apoptosis and cell-cycle control.[4] Heritable mutations in *TP53* are associated with the Li-Fraumeni syndrome, a Mendelian disorder characterized by increased incidence of multiple cancer types.[5] In addition, the majority of epithelial cancers have been shown to carry somatic *TP53* aberrations, mainly mutations of the DNA binding domain of the p53 protein.[6] In cases where *TP53* mutations are not present, p53 function is often abrogated either through loss-of-heterozygosity (LOH) by deletion or methylation of the 17p locus, or through abrogation of the function of p53 downstream effectors. The high frequency of p53 inactivation in human cancers highlights the importance of the p53 tumor suppressor function; for this reason it has been called the "guardian of the genome".[7]

These observations have provided a strong biological rational to the hypothesis that high frequency functional *TP53* polymorphisms may contribute to the population risk of developing common cancers.[4] Most studies have focused on a non-synonymous *TP53* polymorphism in exon 4, where a cytosine (*C*, variant allele) to guanine (*G*) substitution results in the substitution of proline (Pro) for arginine (Arg) at codon 72 of the p53 protein (Arg72Pro, refSNP number: rs1042522).[8] There exists evidence to suggest that the two variant alleles at this locus code for p53 protein isomorphs that differ in their capacities to induce target gene transcription, their interaction with p73 (another tumor suppressor protein), their targeting of the proteasome, and their susceptibility to degradation by human papillomavirus E6 protein.[9-11] These observations have provided the rationale for a large number of genetic association studies investigating rs1042522 as a risk factor for various human malignancies.[12] However, most studies published to date have had small sample sizes, rendering them underpowered to detect small genetic effect sizes, and have often produced contradicting results.

In cervical cancer, where initial evidence suggested a strong protective effect for the Pro-encoding allele[13], a recent meta-analysis of individual patient data failed to identify any association with cancer risk.[14] Intriguingly, a subgroup analysis suggested that the protective effect for the Pro-encoding allele was present in studies that used tumor tissue for genotyping cancer cases. Several lines of evidence indicate that epithelial cancers preferentially retain the Arg-encoding allele. This phenomenon, which represents non-random LOH, could cause directional genotype misclassification affecting only studies using tumor tissue for genotyping cases.[15-19]

We performed a systematic review of studies investigating rs1042522 and the risk for five common epithelial cancers: breast, lung, colorectal, ovarian and endometrial cancer. We used maximum likelihood and Bayesian methods to explore the genetic effect of this polymorphism in common human cancers and to evaluate whether a systematic bias, due to differential genotype misclassification, has affected study results across cancer subtypes.

Methods

Search strategy and eligibility criteria

We considered three common epithelial cancers, which have been extensively investigated in association with the rs1042522 polymorphism: lung, breast and colorectal cancer. In addition, we considered studies of ovarian and endometrial cancer, because these cancers represent a major source of morbidity for women. All five cancers have high public health importance because of their high incidence and associated mortality; in addition, they are all known to frequently exhibit LOH at the *TP53* locus.

We searched the MEDLINE database (through PubMed, last search: March 8th, 2011), to identify genetic association studies of the rs1042522 polymorphism and breast, lung, colorectal, endometrial and ovarian cancer. We used combinations of the following keywords and their synonyms: "*TP53*", "colorectal cancer", "breast cancer", "lung cancer", "endometrial cancer", "ovarian cancer", "Arg72Pro", and "rs1042522". The full search strategy is available in **Appendix Document 1**. These searches were complemented by searches of the Genetic Association Database[20] (GAD, last search: March 8th, 2011) and the Human Genome Epidemiology Network's Literature Finder (HuGe Net, last search: March 8th, 2011).[21] We also used two *TP53*-specific databases that, in addition to collecting data on *TP53* mutations, provide information on *TP53* polymorphisms: the International Agency for Research on Cancer *TP53* database[22-23] (http://www-p53.iarc.fr/) and the p53 Website[24-25] (http://p53.free.fr/). Finally, we hand-searched the reference lists of all identified eligible articles, reviews relevant to our topic and meta-analyses of genetic association studies of the *TP53* polymorphisms.

Studies were considered eligible if they used genotyping methods to determine rs1042522 genotype in patients with breast, lung, colorectal, ovarian or endometrial cancer and in controls with no neoplastic disease. We only considered studies with an analytical epidemiologic design (case control, nested case control or cohort). Studies had to genotype individual samples corresponding to each participant; studies using DNA pooling methods were excluded.[26-27] When multiple studies reported on overlapping patient groups, we included information from the study with the largest number of cancer cases in our analyses. We identified overlap by comparing authors, research centers, recruitment periods, and patient demographic characteristics among otherwise eligible studies. We excluded studies where all individuals had hereditary mutations. For example, we excluded studies enrolling exclusively *BRCA1* or *BRCA2* mutation carriers, studies of individuals with familial adenomatous polyposis coli (FAP) or individuals with hereditary non-polyposis colorectal cancer (HNPCC). We also did not consider family-based studies due to different design and analysis considerations. To limit local literature bias, we only considered studies published in English.[28] Finally, we did not consider editorials, narrative reviews, letters to the editor, or other manuscripts not reporting primary research results.

Data extraction

One reviewer (IJD) screened all abstracts to identify potentially eligible studies and a second reviewer (VV) independently screened abstracts excluded by the first; studies considered potentially eligible by at least one of the reviewers were retrieved and reviewed in full text. One reviewer (IJD) extracted the following information from each eligible study: author, year and journal of publication, number of cases and controls,

participant ethnicity, study design (case-control or cohort), whether cases and controls were matched (for case-control studies), whether controls were sampled from specific disease groups, the genotyping method used (RFLP versus all other methods), whether any genotyping quality control process was employed (defined as repeat genotyping of all or a subset of samples either with the same or an alternative genotyping method), and whether genotyping was performed blind to the case-control status of participants. For our primary comparison of interest, we collected information on the source of genetic material that was used for genotyping cases (cancer tissue versus other DNA source, including buccal swabs, peripheral blood, and saliva). Finally we extracted the rs1042522 genotype distributions in cases and controls. When studies did not report all the required information but instead cited relevant publications, we retrieved and extracted data from them. A second reviewer (VV) verified all extracted information and discrepancies were resolved by consensus, involving a third reviewer (TAT).

Evidence synthesis

From each study we calculated the odds ratio (OR) for the genetic effect of the Pro-encoding allele and its variance under an allele frequency comparison (Pro- versus Arg-encoding allele) to ensure consistency and the inclusion of the maximum number of studies. For each cancer, summary odds ratios and their 95% confidence interval (CI) were calculated using random effects models (DerSimonian-Laird).[29-30] This model incorporates between-study heterogeneity in the summary estimate of the genetic effect and – in the presence of heterogeneity – will result in wider CIs around this estimate, compared to fixed effects models.[31] Between-study heterogeneity was assessed using Cochran's chi-square-based Q statistic and the I^2 index.[32-33] Heterogeneity was considered

statistically significant at $p_Q<0.1$. The I^2 index takes values ranging from 0 to 100%, with higher values suggesting greater inconsistency between studies.

For each of the cancers investigated, subgroup analyses were performed by stratifying studies on the following study-level characteristics: ethnicity of participants (White versus East Asian, excluding other ethnicities), control selection (healthy versus non-cancer disease controls), use of genotyping quality control (yes versus no/not reported), blind genotyping (yes versus no/not reported), use of cancer tissue as the source of genetic material for case genotyping (yes versus no). We also estimated the effect of these study-level covariates on the genetic effect using random effects meta-regression.[34-35] In view of the evidence that using tumor tissue as the source of genotyping material for cases may introduce bias, subgroup and meta-regression analyses were only performed among studies that used more appropriate sources of genotyping material.

We also used a generalized linear mixed effects model meta-analysis approach to evaluate the effect of using cancer tissue as the source of material for case genotyping across cancers. Briefly, we fit a two-level logistic regression model using maximum likelihood to take into account clustering of patients within studies (i.e. the first level accounted for variation within study and the second level for variation between studies) and we modeled the effect of genotyping tumor tissue (for cases) as an interaction term, i.e. a factor modifying the genetic effect across cancers.[36-37] In all such analyses, the genetic effect of rs1042522 was allowed to differ between cancers. Details about the modeling approach are presented in **Appendix Document 3**.

We also implemented the same model using a Bayesian two-level logistic regression model,[38] which enabled us to incorporate prior information on the bias arising from using cancer tissue for genotyping.[39-40] The prior was developed from results of a recent individual-patient data (IPD) meta-analysis of the association between rs1042522 and cervical cancer.[14] As a sensitivity analysis for the effect of using alternative prior specifications, we also evaluated results when using a non-informative prior for the bias effect.

Bayesian models were estimated using Markov Chain Monte Carlo (MCMC). For each model we ran 3 MCMC chains for a total of 100,000 iterations, using a burn in of 10,000. Convergence of the MCMC chains was checked by the Brooks-Gelman-Rubin criteria[41-42] and by inspection of trace plots. We evaluated the fit of the Bayesian models based on the summary Deviance Information Criterion (DIC) as well as inspection of shrinkage plots (graphs of posterior estimates for model parameters plotted along with the prior study estimates). DIC is a deviance measure of goodness of fit equal to the posterior mean of minus twice the log likelihood, penalized by an estimate of the effective number of parameters in the model. The DIC is a Bayesian measure analogous to the Akaike Information Criterion used in classical analysis, but which can also be applied to hierarchical models. It penalizes the likelihood for addition of parameters so that models of different complexity can be appropriately compared.[43]

Sensitivity analyses and assessment of bias

We assessed whether studies producing more precise estimates (larger studies) produced different results compared to studies producing less precise estimates (smaller studies) using the Harbord modification of the Egger regression-based test.[44-45] These

tests are often erroneously referred to as “publication bias” tests, however, the term tests “for small study effects” (i.e., tests of association between precision and effect size) may be more appropriate.[46]

We performed sensitivity analyses using alternative prior distributions for Bayesian analyses, including alternative distributions for the heterogeneity parameter (τ^2). These analyses produced quantitatively similar results (not presented here), suggesting that the Bayesian models were not sensitive to choice of priors. We also explored a Bayesian three-level model (see **Appendix Document 3**), in which patients were considered nested within studies and studies were nested within specific cancer topics.

To explore whether a single study affected estimates of the genetic effect in cancer specific meta-analyses, we repeated the cancer-specific meta-analyses by sequentially dropping one study from the analysis and repeating the calculations. We also repeated the cancer-specific meta-analysis calculations using a fixed effects model (Mantel-Haenszel).[47]

Software

Statistical analyses were carried out in Stata version 11.1/SE (StataCorp, College Station, TX), R version 2.11.0 (R Foundation for Statistical Computing, Vienna, Austria), and WinBugs version 1.4.3 (MRC Biostatistics Unit, Cambridge, UK.). Statistical significance for frequentist hypothesis tests was defined as a two-sided p-value<0.05 with no adjustment for multiple comparisons.[48] For Bayesian analyses we reported the medians and 2.5 and 97.5 percentiles of posterior distributions as 95% credibility intervals (CrI).

Results

Eligible studies

Our searches retrieved a total of 7268 unique citations from Medline, HuGE Net Literature Finder, and the *TP53*-related databases searched. 6888 citations were excluded after screening titles and abstracts and 380 were retrieved and reviewed in full text. Of these, 246 were excluded. The most common reasons for exclusion were inclusion of cases only, assessment of irrelevant genes or polymorphisms, non-cancer conditions or cancers other than breast, lung, colorectal, ovarian or endometrial (a detailed list of reasons for exclusion is presented in **Figure 1**). Overall, 134 articles were considered eligible for this review. **Figure 1** presents the details of the search flow and **Appendix Document 2** presents a list of included studies.

In total, the eligible manuscripts reported on 161 case-control sub-studies (some articles presented data on cases and controls sampled from different populations or reported on multiple cancers). We treated these 161 sub-studies as separate strata ("studies") in our analyses because they pertained to different study bases (typically sampled from different geographic locations or belonging to different ethnicities). Overall, 69 studies investigated the association of rs1042522 with breast cancer (28,513 cases/31,850 controls), 42 with lung cancer (16,743 cases/16,504 controls), 26 with colorectal cancer (7377 cases/10,011 controls), 16 with ovarian cancer (1982 cases/5226 controls) and 8 with endometrial cancer (726 cases/1292 controls). Ninety-seven (60%) studies reported on predominantly white populations, 35 (22%) reported on East Asian populations, 4 (2%) on black populations, 5 (3%) on Hispanic populations, and 20 (12%) on populations of mixed or other ethnicities. **Table 1** presents the characteristics of

eligible studies stratified by the cancer investigated, including details regarding the genotyping methods.

All studies included in our analyses had a case-control design. Seventy-three (45%) of the studies reported matching participants for at least one characteristic; age and sex (in lung and colorectal cancer studies) were the most common matching characteristics. The majority of studies enrolled healthy controls (n=124, 77%) and the remaining (n=37; 23%) studies used non-neoplastic disease controls. Studies were published between 1991 and 2011. The median number of cases was 138 (25^{th} percentile = 78; 75^{th} percentile = 288); studies in breast cancer were larger compared to studies in lung or colorectal cancer. The median number of controls was 211 (25^{th} percentile = 109; 75^{th} percentile=434); breast cancer studies tended also to have larger control groups. Overall the number of participants increased over time; the Spearman test p-value was <0.001 both for the number of cases and controls.

Genotyping methods

Very few studies reported that genotyping was blinded to the case-control status of individuals (n=22, 14%). Genotyping quality control procedures were employed in 64 (40%) of the studies. Genotyping was through RFLP-based methods in most studies (75, 47%). Twenty-two of the studies (14%) used tumor tissue as the source of genotyping material for cases. Specifically, 9 studies on breast cancer, 3 on lung cancer, 6 on colorectal cancer, 2 on ovarian cancer and 2 on endometrial cancer explicitly stated that tumor tissue was used as the source of genotyping material and did not report the use of any technique that would ensure the presence of an adequate amount of normal tissue (such as pathological examination or micro-dissection).

rs1042522 and cancer risk

Random effects meta-analyses using all 161 studies found no significant effect for the Pro allele for breast cancer [OR=0.97 (95% CI, 0.92-1.02)], colorectal cancer [OR=1.05 (95% CI, 0.94-1.17)], ovarian cancer [OR=1.05 (95% CI, 0.93-1.18)] or endometrial cancer [OR=1.02 (95% CI, 0.83-1.26)], but did find some evidence of an increase in the risk of lung cancer [OR=1.07 (95% CI, 1.01-1.14)]. The meta-analyses of breast, lung and colorectal cancer found substantial between-study heterogeneity: the p-value of the Q-statistic was <0.001 in all three cases and the I^2 index values were 59% (95% CI, 47%-69%), 58% (95% CI, 41%-70%) and 75% (95% CI, 63%-83%) for breast, lung and colorectal cancer, respectively. Heterogeneity was less pronounced in the ovarian and endometrial cancer meta-analyses. The Q-statistic p-values were 0.04 and 0.11 and the I^2 values were 42% (95% CI, 0%-68%) and 40% (95% CI, 0%-74%), respectively.

Figure 2 presents the summary estimates for the effect of rs1042522 on cancer risk, stratified by cancer and the tissue used for case genotyping. The studies using tumor tissue for case genotyping had summary ORs of 0.78 (95% CI, 0.58-1.04), 0.70 (95% CI, 0.43-1.12), 0.88 (95% CI, 0.60-1.27), 0.47 (95% CI, 0.30-0.74) and 0.73 (95% CI, 0.38-1.43) for breast, lung, colorectal ovarian and endometrial cancer, respectively. In contrast, subgroup analyses suggested that studies using more appropriate sources of material for case genotyping demonstrated no significant association for breast, colorectal, or endometrial cancer, OR=0.98 (95% CI, 0.94-1.03), OR=1.09 (95% CI, 0.97-1.22) and OR= 1.10 (95% CI, 0.91-1.33) respectively, and even suggested an increased risk for lung, OR=1.09 (95% CI, 1.03-1.16; p=0.003), and ovarian cancer, OR = 1.10 (95% CI, 1.01-1.19; p=0.031). It is evident that studies using tumor tissue suggest a protective

effect for the Pro allele (statistically significant only for ovarian cancer; p=0.001) whereas studies using other sources of genotyping material either found no effect (for breast, colorectal and endometrial cancer) or even indicated an increased risk for the Pro allele (for lung and ovarian cancer). Forest plots for each cancer separately are presented in **Appendix Figures 1 to 5**.

Based on meta-regression analyses for each cancer, the relative OR comparing studies using tumor tissue with those using other sources of genotyping material was 0.82 (95% CI, 0.65-1.03) for breast cancer, 0.67 (95% CI, 0.47-0.96) for lung cancer, 0.81 (95% CI, 0.58-1.12) for colorectal cancer, 0.43 (95% CI, 0.27-0.67) for ovarian cancer and 0.69 (95% CI, 0.41-1.17) for endometrial cancer, in all cases suggesting that studies using tumor tissue tend to produce lower ORs for the Pro allele, i.e., that these studies systematically underestimate the genetic effect of the Pro allele.

Assessment of the bias effect across cancers

To further explore and quantify the bias arising from using cancer tissue as the source of genotyping materials, we used a multilevel model that allowed us to borrow strength by combining information across cancers. The details of the modeling approach are presented in **Appendix Document 3**; it is important to note that although the bias effect is assumed to be common across cancers, the genetic effect (adjusted for any genotyping bias) of the Pro allele is allowed to differ across cancers. Using a two-level mixed effects logistic regression model, we estimated this common bias effect to have a relative OR = 0.78 (95% CI, 0.69-0.90) with a p-value<0.001. This result suggests that studies using tumor tissue will underestimate the genetic effect of the Pro-allele by approximately 20%. **Figure 3** presents the summary estimates for the effect of rs1042522

on cancer risk, stratified by cancer and the tissue used for case genotyping from the multilevel model fit using maximum likelihood.

To allow direct probabilistic statements to be made based on the model incorporating information from all cancers, we also used a Bayesian approach with an informative prior. The prior was derived from the subgroup estimates (by source of DNA for genotyping) from an individual patient data meta-analysis of rs1042522 and cervical cancer risk.[14] In this analysis the bias effect was estimated to be 0.78 (95% CrI, 0.70-0.88). **Table 4** presents the meta-analysis results for this model. The posterior probability that use of tissue for genotyping cases is associated with an underestimation of the Pro allele's effect, i.e., the probability that using cancer tissue as a source of genotyping material leads to underestimation of the "true" genetic effect of the Pro allele, was estimated to be higher than 99%. Using a non-informative prior, the results were very similar: the bias was estimated to have a relative OR=0.79 (95% CrI 0.69-0.89) and the posterior probability of the effect being lower than 1 was estimated to be >99% (**Table 5**). As expected, estimates of the bias effect from models using non-informative priors are very close to those from maximum likelihood estimation. **Appendix Figure 6** presents the posterior distributions for the summary OR of each cancer and the bias effect. For Bayesian analyses, model convergence was adequate as was evident from the Gelman-Rubin statistics for each model parameter (all values <1.01), the multivariate Gelman-Rubin statistic for each model (all values<1.01) and the graphical assessment of trace plots. **Appendix Figure 7** presents the trace plot corresponding to the 2-level Bayesian model discussed above.

Other modifiers of the genetic effect

To assess whether other covariates modified the genetic effect of the Pro allele, we used subgroup and univariable meta-regression analyses, separately for each cancer of interest, for the following study-level characteristics: participant ethnicity, study design, control selection methods, the specific genotyping method used, whether genotyping quality control procedures were used and whether investigators were blind to the case/control status of participants. Given the potential for bias in studies using tumor tissue for genotyping cases, such studies were excluded from these analyses.

Overall, the results of stratified and regression analyses did not indicate any significant modification of the genetic effect by the majority of covariates assessed; **Appendix Tables 1-6** present the results of subgroup and meta-regression analyses results. Potential exceptions were ethnicity and use of RFLP genotyping methods in studies of colorectal cancer and whether the genotype frequencies of the control group were in HWE in studies of breast, colorectal and endometrial cancer. Specifically, in studies of colorectal cancer the genetic effect of the Pro-allele appeared to be stronger in studies of East Asian participants [relative OR = 1.21 (95% CI, 1.03-1.44; p-value=0.023)], studies using RFLP genotyping methods [relative OR = 1.31 (95% CI, 1.04-1.65; p-value=0.023)], and studies where the control group genotype frequencies were not consistent with HWE [relative OR=2.93 (95% CI, 1.79-4.80; p-value <0.001)]. In studies of breast and endometrial cancer the genetic effect was lower in studies where the genotype frequencies in the control group deviated from HWE [relative OR = 0.85 (95% CI, 0.73-0.99; p-value =0.042) and 0.60 (95% CI, 0.38-0.94; p-value = 0.026), respectively].

Sensitivity analyses and assessment of bias

There was no evidence of a systematic difference in the effect sizes reported in smaller versus larger studies for any of the cancers of interest. The Harbord test p-values were 0.184, 0.071, 0.290, 0.985 and 0.680 for breast, lung, colorectal, ovarian and endometrial cancer, respectively. Inclusion of studies that used tissue as the source of genotyping material for cases changed the results for breast and lung cancer towards a suggestion of a systematic difference between smaller and larger studies (p=0.049 and 0.017, respectively). This finding possibly reflects the fact that studies using tumor tissue as the source of genotyping material for cases were generally smaller compared to studies using more appropriate DNA sources.

There was no evidence that the first published study assessing the association of rs1042522 with any of the cancers we evaluated produced more extreme results compared to all subsequent studies. The relative ORs comparing the first with all subsequently published studies on each cancer were 0.92 (95% CI, 0.56-1.51), 0.99 (95% CI, 0.52-1.87), 0.72 (95% CI, 0.23-2.23), 1.03 (95% CI, 0.56-1.91) and 1.33 (95% CI, 0.15-12.06) for breast lung, colorectal, ovarian and endometrial cancer, respectively. The corresponding interaction p-values comparing the first with all subsequent published studies were 0.747, 0.968, 0.549, 0.913 and 0.738 for breast, lung, colorectal, ovarian and endometrial cancer, respectively.

Sensitivity analyses using alternative prior distributions for Bayesian analyses, including alternative distributions for the heterogeneity parameter, produced results similar to the main analyses, highlighting the robustness of our results to model specification. We also explored fitting a Bayesian three-level model (see **Appendix Document 3**), in which patients were considered nested within studies and the specific

cancer topic was considered as an additional level. Using the informative prior from the cervical cancer meta-analysis suggested that the posterior probability that use of tumor tissue for genotyping cases was associated with an underestimation of the Pro-encoding allele's genetic effect was again higher than 99% (**Table 6**). Using a non-informative prior, this model gave a probability of 97.8% that the bias effect leads to underestimation of the genetic effect of the Pro allele (**Table 7**).

Finally, among studies using appropriate sources of genotyping material, leave-one-out meta-analyses and analyses using a fixed effects model produced similar inferences with our main analyses (data not shown).

Discussion

TP53 rs1042522 is one of the most commonly investigated variants in cancer genetic epidemiology.[12] This systematic review provides compelling evidence that this polymorphism is unlikely to be a risk factor for breast, colorectal, ovarian or endometrial cancer and suggests that the Pro allele at this locus may cause a small increase in the risk of lung cancer. More importantly, our work provides evidence that some of the findings of genetic association studies of this variant may have been driven by differential genotype misclassification, in cases where tumor tissue was used as the source of genotyping material for cancer cases. We used different analytic approaches that allowed us to borrow strength across cancers and increase the precision of the estimate of this bias effect. Based on the published data on all 5 cancers we evaluated, use of tumor tissue appears to lead to an underestimation of the genetic effect by approximately 20%. The probability that using tumor tissue actually biases estimates of the genetic effect downward was higher than 99% when we incorporated prior evidence from a recent meta-analysis of the same polymorphism in cervical cancer and higher than 96% in all analyses. The fact that studies using cancer tissue are susceptible to bias is both biologically plausible and has empirical support from studies in other cancers.

We hypothesize that the misclassification arises due to preferential LOH of the Pro-allele in heterozygous individuals.[15-16, 18-19] A substantial body of research, using matched samples of peripheral blood (or other sources of genotyping material, including buccal swabs or saliva) and tumor tissue from cancer patients, demonstrates that LOH at the *TP53* locus is non-random and preferentially involves the Pro-allele. This phenomenon has been documented in the cancers considered in this review (breast, lung,

colorectal, ovarian, endometrial), as well as other cancer types, such as cervical, urothelial, and head and neck cancer.[16-17, 19, 49-52] The biological mechanisms underlying this phenomenon are largely uncertain, but it appears that the occurrence of *TP53* mutations is more common on the DNA strand carrying the Arg-encoding allele.[15, 53] This suggests that preferential retention of the Arg-allele may be the result of an anti-apoptotic advantage conferred by the *TP53* mutation.

From an epidemiological standpoint, our findings are consistent with a large individual-patient data meta-analysis of 49 studies (7946 cases and 7888 controls) investigating the association between rs1042522 and cervical cancer risk.[14] The meta-analysis concluded: “excess risks were most likely not due to clinical or biological factors, but to errors in study methods”. Specifically, studies in which the genotype of cases was determined from white blood cells produced null results; in contrast, studies where genotype was determined using tumor tissue suggested a protective effect of the Pro allele. Our Bayesian analyses incorporated these findings in the form of an informative prior distribution and suggested that the bias towards a protective effect of the Pro allele operates across cancers.

Our findings suggest that use of tumor-derived DNA in genetic association studies should be avoided because it can appreciably bias the results of genetic association studies. This may be particularly true for variants in regions where LOH is known to occur, and may present an important concern for pharmacogenetic studies where the only available source of genotyping material is often tumor tissue. When the interest is in identifying germline (i.e. non-somatic) variants potentially associated with treatment outcomes, it may be prudent to obtain paired normal-tumor samples at least from a

random sample of participants, in order to establish the extent of LOH as well as to gauge the effect it may have on the overall study results. Currently, such information is available only for a minority of genes, based on studies enrolling small numbers of participants.

When analysis was limited to studies using appropriate sources of material for genotyping cases, presence of the Pro allele with lung cancer was positively associated with lung cancer, OR=1.09 (95% CI, 1,02-1.15; p=0.006, based on the multilevel model fit using maximum likelihood). This finding is consistent with previously published meta-analyses on this association.[54-55] However, given the small effect size we observed (OR=1.09) and the fact that in Bayesian analyses the 95% CrI for the lung cancer OR was very close to one, further studies may be necessary to confirm our results. Furthermore, among lung cancer studies, there was some indication of a systematic difference between smaller and larger studies. Given the substantial between-study heterogeneity (I^2=58%) for this association, both "true" between-study differences and publication bias are tenable explanations for this discrepancy.[46]

We found no indication of temporal changes in the summary effect size of each meta-analysis or indications of any association between the genetic effect and patient ethnicity or aspects of study design. We also explored the interaction between several study-level covariates relevant to the genotyping procedures employed and the magnitude of the genetic effect. There was little evidence that blinding investigators to the case/control status of samples, use of quality control or specific genotyping methods influenced the estimate of the genetic effect from individual studies.

Aspects of the genotyping process, such as the selection of samples, their procurement and the specific genotyping method used, have received relatively little attention compared to other aspects of study design.[56-57] However, these issues have come to the forefront in the era of genome-wide association studies because, with increasing sample sizes, even small effects from genotyping errors can reach statistical significance and distort study results.[58] The present work identifies a source of bias that appears to have affected the results of candidate gene studies. The extent to which similar issues may be present for other genetic loci is unknown.

Our work has several strengths: first, we used a comprehensive search strategy across multiple databases and identified a large number of informative studies for our analyses. Second, we believe there is little potential of selective reporting bias to have affected our results as only a few otherwise eligible studies were excluded due to inability to extract genotype data. Most importantly, we were able to extract information on the source of genotyping material from all studies (i.e. the relevant information was always reported), indicating that misclassification of this variable in our dataset is unlikely. Third, our finding that use of cancer tissue as a source of genotyping material may lead to underestimation of the Pro allele effect was consistent across cancers and robust to the use of alternative analytical approaches.

Several limitations need to be considered when interpreting our results. Our analysis was based on published data, preventing any statistical adjustment for individual level factors that might modify the genetic effect, such as gender (mostly for lung and colorectal cancer), age or smoking (a major determinant of lung cancer risk). However, we note that the potential for confounding bias is limited due to the random assortment of

alleles at meiosis (Mendelian randomization).[59] Furthermore, we could not assess gene-environment interactions because data on potential exposures of interest were unavailable from most studies. Nonetheless, it is unlikely that such factors would also influence the estimate of the bias effect. Finally, our findings should primarily be viewed as hypothesis generating; large studies reporting on paired samples of normal and tumor tissue obtained from the same patients are needed to definitively confirm our observations.

In conclusion, our analyses demonstrate that *TP53* rs1042522 is unlikely to be associated with breast, colorectal or endometrial cancer; however a weak association with lung and ovarian cancer may exist. Across cancer types, there is compelling evidence that using genetic material obtained from tumor tissue to genotype cases can bias the estimate of the genetic effect, leading to a 25% underestimation of the Pro allele's effect; the probability that bias is towards a protective effect for the Pro allele was at least 95% in our analyses. This finding, along with laboratory evidence indicating that LOH at the *TP53* locus in many epithelial cancers is non-random, suggests that studies using tumor as the source of genotyping material for cases have been affected by differential genotype misclassification. Future studies, particularly when the genetic loci of interest are known to exhibit LOH, should avoid the use of tumor tissue as the primary source of genetic material.

Tables

Table 1

Characteristics of eligible studies.[a]

Study characteristics	Breast cancer (69 studies)	Lung cancer (42 studies)	Colorectal cancer (26 studies)	Ovarian cancer (16 studies)	Endometrial cancer (8 studies)	All studies (161 studies)
Number of cases/ controls	28513/ 31850	16743/ 16504	7377/ 10011	1982/ 5226	726/ 1292	55341/ 64883
Median number of cases	175	147	121	109	94	138
25^{th} – 75^{th} percentile	94-468	91-307	76-345	48-193	43-118	78-288
Median number of controls	218	176	220	281	78	211
25^{th} – 75^{th} percentile	109-500	133-379	140-347	74-446	31-310	109-434
Ethnicity						
White	46 (67)	16 (38)	16 (62)	13 (81)	6 (75)	97 (60)
East Asians	10 (14)	13 (31)	8 (31)	2 (13)	2 (25)	35 (22)
Black	1 (1)	2 (5)	0 (0)	1 (6)	0	4 (2)
Latinos	1 (1)	3 (7)	1 (4)	0	0	5 (3)
Others/ mixed/ NR	11 (16)	8 (19)	1 (4)	0	0	20 (12)
Control selection						
Healthy	61 (88)	23 (55)	19 (73)	15 (94)	6 (75)	124 (77)
Diseased	8 (12)	19 (45)	7 (27)	1 (6)	2 (25)	37 (23)
Matched Controls						
Yes	32 (46)	23 (55)	10 (38)	6 (37)	2 (25)	73 (45)
No/ not applicable	37 (54)	19 (45)	16 (62)	10 (63)	6 (75)	88 (55)
Blinding to case-control status						
Yes	8 (12)	9 (21)	3 (12)	2 (12)	0	22 (14)
No/NR	61 (88)	33 (79)	23 (88)	14 (88)	8 (100)	139 (86)
Use of genotyping QC						
Yes	25 (36)	17 (40)	11 (42)	9 (56)	2 (25)	64 (40)
No/NR	44 (64)	25 (60)	15 (57)	7 (44)	6 (75)	97 (60)
Genotyping methods						
RFLP	33 (48)	24 (57)	13 (50)	2 (12)	3 (36)	75 (47)
Other methods	36 (52)	18 (43)	13 (50)	14 (88)	5 (63)	86 (53)
HWE						
Compliant	54 (78)	36 (86)	24 (92)	13 (81)	7 (88)	134 (83)
In violation	15 (22)	6 (14)	2 (8)	3 (19)	1 (12)	27 (17)
Use of tumor tissue for genotyping[b]						
Yes	9 (13)	3 (7)	6 (23)	2 (12)	2 (25)	22 (14)
No	60 (87)	39 (93)	20 (77)	14 (88)	6 (75)	139 (86)

[a] Unless otherwise stated, numbers indicate the number of studies and % of studies that had each characteristic of interest.

[b] In all cases, studies provided adequate data to evaluate the source of DNA for case genotyping.

HWE = Hardy-Weinberg equilibrium; NR = not reported; QC = quality control; RFLP = restriction-fragment length polymorphism.

Percentages have been rounded to the nearest integer. Percentages may not sum up to one due to rounding.

Table 2

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a random effects model (DerSimonian-Laird), separately for each cancer.

Subgroup	Breast cancer	Lung cancer	Colorectal cancer	Ovarian cancer	Endometrial cancer
			OR (95% CI); p_Q, I^2		
All studies	0.97 (0.92-1.02); <0.001, 59%	1.07 (1.01-1.14); <0.001, 58%	1.05 (0.94-1.17); <0.001, 75%	1.05 (0.93-1.18); 0.042, 42%	1.02 (0.83-1.26); 0.110, 40%
Studies with somatic DNA used for case genotyping	0.98 (0.94-1.03); <0.001, 56%	1.09 (1.03-1.16); <0.001, 54%	1.09 (0.97-1.22); <0.001, 75%	1.10 (1.01-1.20); 0.534, 0%	1.10 (0.91-1.33); 0.265, 22%
Studies using tumor tissue for case genotyping	0.78 (0.58-1.04); <0.001, 68%	0.70 (0.43-1.12); 0.06, 65%	0.88 (0.60-1.27); 0.002, 74%	0.47 (0.30-0.74); 0.549, 0%	0.73 (0.38-1.43); 0.103, 62%

CI = confidence interval; OR = odds ratio; p_Q = p-value of Cochran's Q statistic.

Table 3

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a 2-level mixed effects logistic regression (maximum likelihood) model.

Cancer	Studies using appropriate DNA sources OR (95% CI)	p-value	Studies using tumor tissue OR (95% CI)	p-value
Breast cancer	0.98 (0.94-1.03)	0.484	0.77 (0.68-0.88)	<0.001
Lung cancer	1.09 (1.02-1.15)	0.006	0.86 (0.74-0.99)	0.025
Colorectal cancer	1.10 (0.97-1.25)	0.155	0.86 (0.73-1.01)	0.074
Ovarian cancer	1.05 (0.94-1.18)	0.362	0.83 (0.70-0.98)	0.025
Endometrial cancer	1.09 (0.91-1.30)	0.363	0.85 (0.69-1.05)	0.133
Bias effect (across cancers)	0.78 (0.69-0.90), p<0.001			

CI = confidence interval; OR = odds ratio.

Table 4

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a 2-level mixed effects logistic regression (Bayesian implementation) model (informative prior).

Cancer	Studies using appropriate DNA sources OR (95% CrI)	Studies using tumor tissue OR (95% CrI)
Breast cancer	0.99 (0.94-1.03)	0.77 (0.68-0.87)
Lung cancer	1.08 (1.00-1.16)	0.85 (0.74-0.97)
Colorectal cancer	1.09 (0.99-1.20)	0.86 (0.75-0.98)
Ovarian cancer	1.04 (0.91-1.19)	0.82 (0.68-0.97)
Endometrial cancer	1.08 (0.88-1.32)	0.84 (0.67-1.06)
Bias effect (across cancers)	0.78 (0.70-0.88)	
*Probability bias <0	>0.999	

CrI = credibility interval; OR = odds ratio.

*This effectively expresses the probability that use of tumor tissue as the source of genotyping material for cases leads to underestimation.

Table 5

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a 2-level mixed effects logistic regression (Bayesian implementation) model (non-informative prior).

Cancer	Studies using appropriate DNA sources OR (95% CrI)	Studies using tumor tissue OR (95% CrI)
Breast cancer	0.99 (0.94-1.04)	0.78 (0.68-0.88)
Lung cancer	1.08 (1.00-1.17)	0.85 (0.74-0.98)
Colorectal cancer	1.09 (0.99-1.20)	0.86 (0.75-0.99)
Ovarian cancer	1.05(0.92-1.19)	0.83 (0.69-0.99)
Endometrial cancer	1.08 (0.88-1.32)	0.85 (0.68-1.06)
Bias effect (across cancers)	0.79 (0.69-0.89)	
*Probability bias <0	>0.999	

CrI = credibility interval; OR = odds ratio.

*This effectively expresses the probability that use of tumor tissue as the source of genotyping material for cases leads to underestimation of the genetic effect of the Pro-allele.

Table 6

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a 3-level mixed effects logistic regression (Bayesian implementation) model (informative prior).

Cancer	Studies using appropriate DNA sources OR (95% CrI)	Studies using tumor tissue OR (95% CrI)
Breast cancer	1.07 (0.98-1.20)	0.85 (0.69-1.08)
Lung cancer	1.08 (1.01-1.14)	0.74 (0.56-0.92)
Colorectal cancer	1.00 (0.94-1.05)	0.81 (0.69-0.97)
Ovarian cancer	1.06 (0.98-1.16)	0.68 (0.41-0.87)
Endometrial cancer	1.06 (0.93-1.25)	0.75 (0.52-1.03)
Bias effect (across cancers)	0.76 (0.63-0.90)	
[*]Probability bias <0	0.997	

CrI = credibility interval; OR = odds ratio.

[*]This effectively expresses the probability that use of tumor tissue as the source of genotyping material for cases leads to underestimation of the genetic effect of the Pro-allele.

Table 7

Meta-analysis results for breast, lung, colorectal, ovarian, and endometrial cancer using a 3-level mixed effects logistic regression (Bayesian implementation) model (non-informative prior).

Cancer	Studies using appropriate DNA sources OR (95% CrI)	Studies using tumor tissue OR (95% CrI)
Breast cancer	1.07 (0.98-1.20)	0.85 (0.68-1.10)
Lung cancer	1.08 (1.02-1.15)	0.73 (0.54-0.92)
Colorectal cancer	0.99 (0.94-1.05)	0.81 (0.68-0.98)
Ovarian cancer	1.06 (0.98-1.17)	0.65 (0.39-0.87)
Endometrial cancer	1.06 (0.93-1.27)	0.75 (0.50-1.02)
Bias effect (across cancers)	0.75 (0.51-0.99)	
[*]Probability bias <0	0.978	

CrI = credibility interval; OR = odds ratio.

[*]This effectively expresses the probability that use of tumor tissue as the source of genotyping material for cases leads to underestimation of the genetic effect of the Pro-allele.

Figures

Figure 1

Search strategy and study eligibility flow.

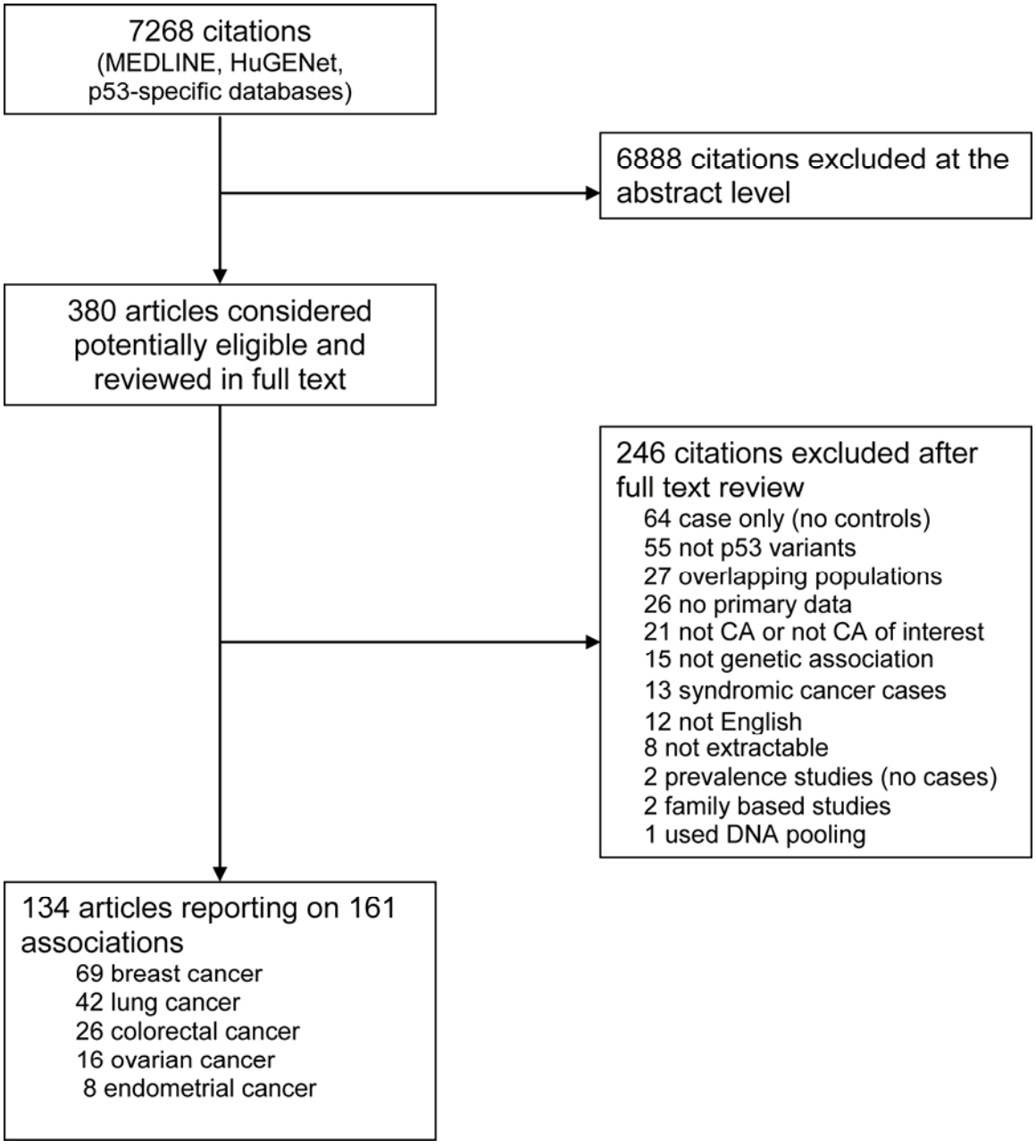
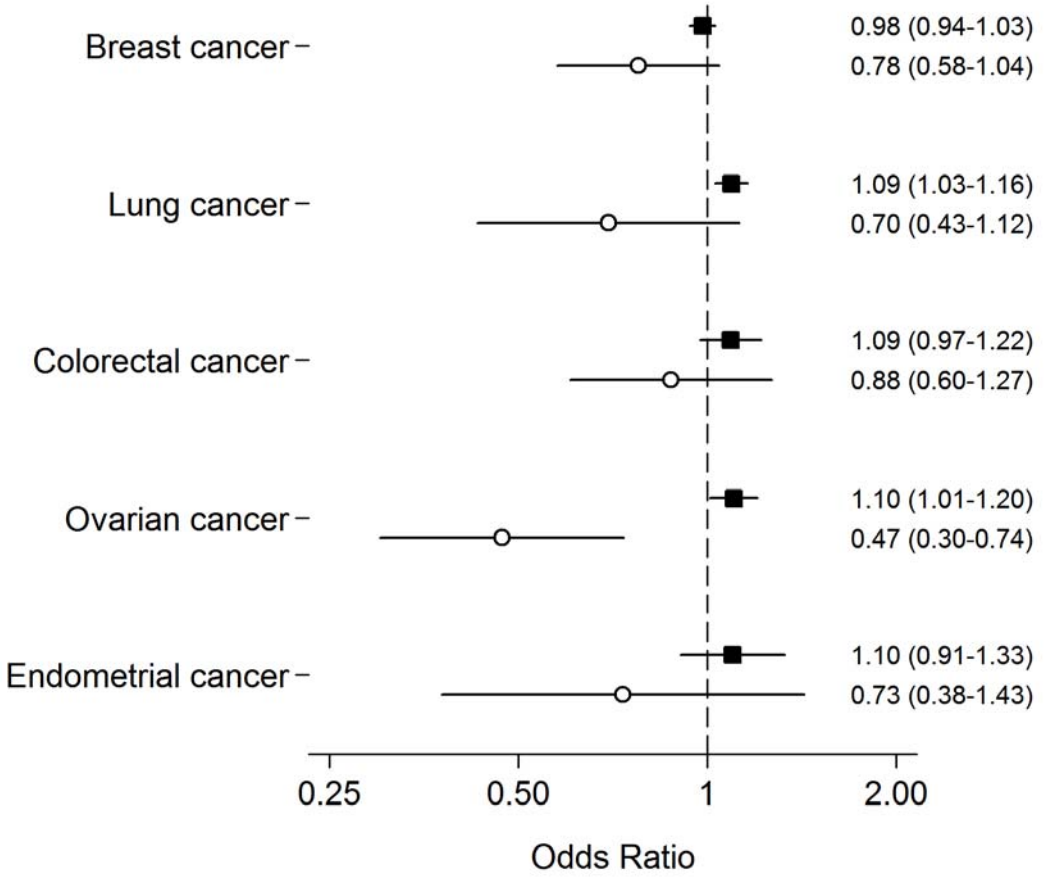


Figure 2

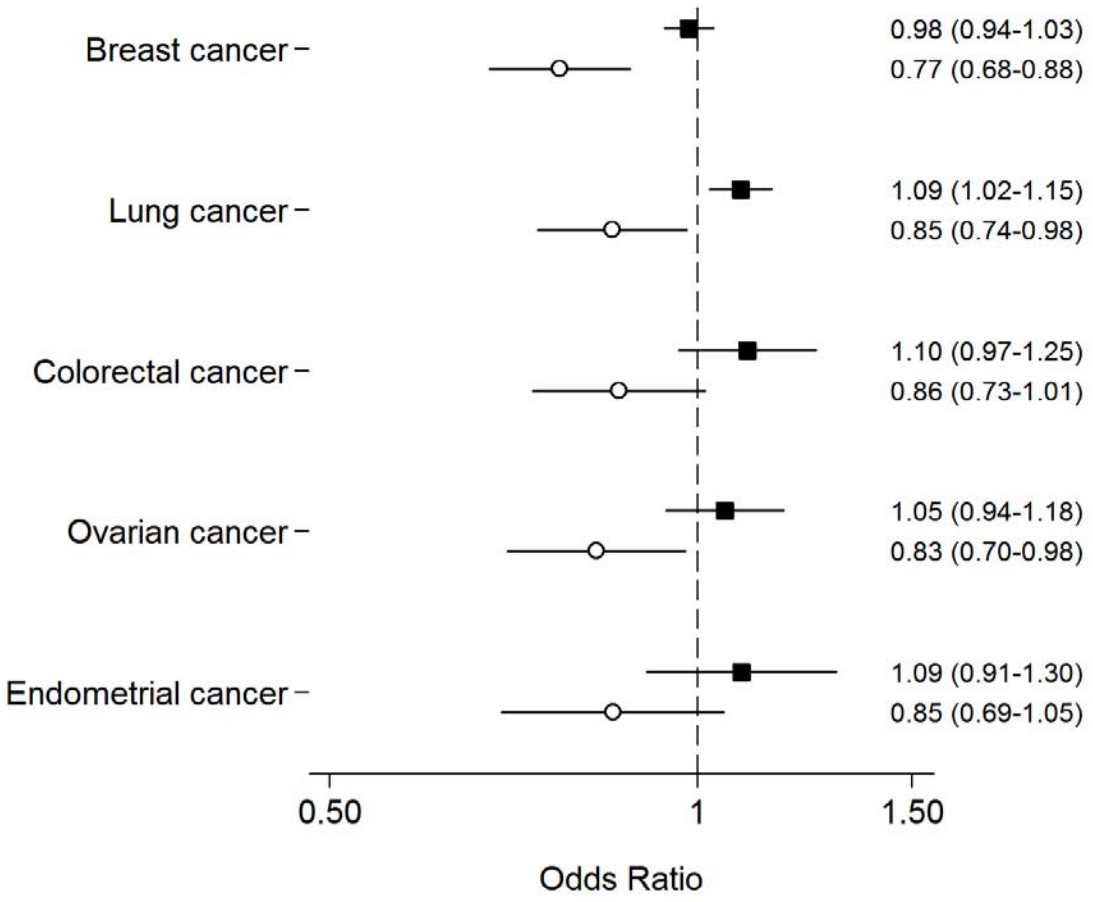
Summary meta-analysis results for breast, lung, colorectal, ovarian and endometrial cancer, stratified by the source of genotyping material for cases.



This forest plot summarizes the results of meta-analyses for each of the cancers of interest, stratified by the source of material used for case genotyping. Results are derived from random effects meta-analysis for each subset of studies (DerSimonian-Laird model). Point estimates from studies using appropriate DNA sources are shown as black squares and point estimates from studies using tumor tissue as the source of genotyping material are shown as white circles. Extending lines represent the 95% confidence interval of each estimate.

Figure 3

Summary meta-analysis results for breast, lung, colorectal, ovarian and colorectal cancer stratified by the source of genotyping material for cases. Estimates were derived from a 2-level logistic regression model fitted using maximum likelihood.



This forest plot summarizes the results of meta-analyses for each of the cancers of interest, stratified by the source of material used for case genotyping. Results are derived from a 2-level mixed effects logistic regression model that incorporated evidence from all cancers to estimate the bias effect. Point estimates from studies using appropriate DNA sources are shown as black squares and point estimates from studies using tumor tissue as the source of genotyping material are shown as white circles. Extending lines represent the 95% confidence interval of each estimate. Note the change in scale compared to Figure 2; summary point estimates for each subgroup from the multilevel model are in most cases similar to the cancer-specific analyses (Figure 2), however precision is greatly increased.

Appendix documents

Appendix Document 1: Search strategies.

MEDLINE

(tp53[All Fields] OR p53[All Fields] OR p-53[All Fields] OR tp-53[All Fields]) AND ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields] OR carcinoma* OR cancer OR cancer? OR neoplasm* OR adenocarcinoma*)
AND (("polymorphism, genetic"[MeSH Terms] OR ("polymorphism"[All Fields] AND "genetic"[All Fields]) OR "genetic polymorphism"[All Fields] OR "polymorphism"[All Fields]) OR arg72pro[All Fields] OR arg72[All Fields] OR pro72[All Fields] OR pro72arg[All Fields] OR rs1042522[All Fields] OR variant* OR variati* OR "haplotypes"[MeSH Terms] OR "haplotypes"[All Fields] OR "haplotype"[All Fields] OR polymorph*)

Human Genome Epidemiology Literature Finder

Database: www.hugenavigator.net
Search criteria: All publications>>TP53, TP53I3, TP53BP2, TP53RK, TP53INP1, TP53AIP1[Gene]>>Mammary Neoplasms, Invasive Ductal Breast Carcinoma, Carcinoma, Endometrioid, Noninfiltrating Intraductal Carcinoma, Carcinoma, Non-Small-Cell Lung, Colonic Neoplasms, Colorectal Neoplasms, Hereditary Nonpolyposis Colorectal Neoplasms, Endometrial Neoplasms, Neoplasm of lung (disorder), Neoplasms, Hormone-Dependent, ovarian neoplasm, Rectal Neoplasms, Small cell carcinoma of lung[Mesh]

<u>*Note:*</u> In addition to *TP53*, we searched for genes belonging to the p53 pathway to increase the sensitivity of the search (in studies of these genes *TP53* polymorphisms are also often investigated).

Other databases (*International Agency for Research on Cancer TP53 database, the p53 Website, Genetic Association Database*)

These databases provide annotated bibliography lists; as such, no specific search strategy was constructed.

Appendix Document 2: List of included studies.

1. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. J Natl Cancer Inst. 2006; **98**(19): 1382-96.
2. Agorastos T, Masouridou S, Lambropoulos AF, Chrisafi S, Miliaras D, Pantazis K, et al. P53 codon 72 polymorphism and correlation with ovarian and endometrial cancer in Greek women. Eur J Cancer Prev. 2004; **13**(4): 277-80.
3. Akkiprik M, Sonmez O, Gulluoglu BM, Caglar HB, Kaya H, Demirkalem P, et al. Analysis of p53 gene polymorphisms and protein over-expression in patients with breast cancer. Pathol Oncol Res. 2009; **15**(3): 359-68.
4. Alawadi S, Ghabreau L, Alsaleh M, Abdulaziz Z, Rafeek M, Akil N, et al. P53 gene polymorphisms and breast cancer risk in Arab women. Med Oncol. 2010.
5. Aoki MN, da Silva do Amaral Herrera AC, Amarante MK, do Val Carneiro JL, Fungaro MH, Watanabe MA. CCR5 and p53 codon 72 gene polymorphisms: implications in breast cancer development. Int J Mol Med. 2009; **23**(3): 429-35.
6. Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, Attia J, et al. Polymorphisms in TP53 and MDM2 combined are associated with high grade endometrial cancer. Gynecol Oncol. 2009; **113**(1): 109-14.
7. Baynes C, Healey CS, Pooley KA, Scollen S, Luben RN, Thompson DJ, et al. Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. Breast Cancer Res. 2007; **9**(2): R27.
8. Birgander R, Sjalander A, Rannug A, Alexandrie AK, Sundberg MI, Seidegard J, et al. P53 polymorphisms and haplotypes in lung cancer. Carcinogenesis. 1995; **16**(9): 2233-6.
9. Biros E, Kalina I, Biros I, Kohut A, Bogyiova E, Salagovic J, et al. Polymorphism of the p53 gene within the codon 72 in lung cancer patients. Neoplasma. 2001; **48**(5): 407-11.
10. Bisof V, Salihovic MP, Narancic NS, Skaric-Juric T, Jakic-Razumovic J, Janicijevic B, et al. TP53 gene polymorphisms and breast cancer in Croatian women: a pilot study. Eur J Gynaecol Oncol. 2010; **31**(5): 539-44.
11. Buller RE, Sood A, Fullenkamp C, Sorosky J, Powills K, Anderson B. The influence of the p53 codon 72 polymorphism on ovarian carcinogenesis and prognosis. Cancer Gene Ther. 1997; **4**(4): 239-45.
12. Buyru N, Altinisik J, Demokan S, Dalay N. p53 genotypes and haplotypes associated with risk of breast cancer. Cancer Detect Prev. 2007; **31**(3): 207-13.
13. Buyru N, Altinisik J, Isin M, Dalay N. p53 codon 72 polymorphism and HPV status in lung cancer. Med Sci Monit. 2008; **14**(9): CR493-7.
14. Buyru N, Tigli H, Dalay N. P53 codon 72 polymorphism in breast cancer. Oncol Rep. 2003; **10**(3): 711-4.
15. Caceres DD, Quinones LA, Schroeder JC, Gil LD, Irarrazabal CE. Association between p53 codon 72 genetic polymorphism and tobacco use and lung cancer risk. Lung. 2009; **187**(2): 110-5.
16. Cao Z, Song JH, Park YK, Maeng EJ, Nam SW, Lee JY, et al. The p53 codon 72 polymorphism and susceptibility to colorectal cancer in Korean patients. Neoplasma. 2009; **56**(2): 114-8.

17. Cavallone L, Arcand SL, Maugard C, Ghadirian P, Mes-Masson AM, Provencher D, et al. Haplotype analysis of TP53 polymorphisms, Arg72Pro and Ins16, in BRCA1 and BRCA2 mutation carriers of French Canadian descent. BMC Cancer. 2008; **8**: 96.
18. Chua HW, Ng D, Choo S, Lum SS, Li H, Soh LY, et al. Effect of MDM2 SNP309 and p53 codon 72 polymorphisms on lung cancer risk and survival among non-smoking Chinese women in Singapore. BMC Cancer. 2010; **10**: 88.
19. Costa S, Pinto D, Pereira D, Rodrigues H, Cameselle-Teijeiro J, Medeiros R, et al. Importance of TP53 codon 72 and intron 3 duplication 16bp polymorphisms in prediction of susceptibility on breast cancer. BMC Cancer. 2008; **8**: 32.
20. Cox DG, Deer D, Guo Q, Tworoger SS, Hankinson SE, Hunter DJ, et al. The p53 Arg72Pro and MDM2 -309 polymorphisms and risk of breast cancer in the nurses' health studies. Cancer Causes Control. 2007; **18**(6): 621-5.
21. Csejtei A, Tibold A, Varga Z, Koltai K, Ember A, Orsos Z, et al. GSTM, GSTT and p53 polymorphisms as modifiers of clinical outcome in colorectal cancer. Anticancer Res. 2008; **28**(3B): 1917-22.
22. Dakouras A, Nikiteas N, Papadakis E, Perakis M, Valis D, Rallis G, et al. P53Arg72 homozygosity and its increased incidence in left-sided sporadic colorectal adenocarcinomas, in a Greek-Caucasian population. Anticancer Res. 2008; **28**(2A): 1039-43.
23. Damin AP, Frazzon AP, Damin DC, Roehe A, Hermes V, Zettler C, et al. Evidence for an association of TP53 codon 72 polymorphism with breast cancer risk. Cancer Detect Prev. 2006; **30**(6): 523-9.
24. De Vecchi G, Verderio P, Pizzamiglio S, Manoukian S, Bernard L, Pensotti V, et al. The p53 Arg72Pro and Ins16bp polymorphisms and their haplotypes are not associated with breast cancer risk in BRCA-mutation negative familial cases. Cancer Detect Prev. 2008; **32**(2): 140-3.
25. Denisov EV, Cherdyntseva NV, Litvyakov NV, Slonimskaya EM, Malinovskaya EA, Voevoda MI, et al. TP53 mutations and Arg72Pro polymorphism in breast cancers. Cancer Genet Cytogenet. 2009; **192**(2): 93-5.
26. Ebner F, Schremmer-Danninger E, Rehbock J. The role of TP53 and p21 gene polymorphisms in breast cancer biology in a well specified and characterized German cohort. J Cancer Res Clin Oncol. 2010; **136**(9): 1369-75.
27. Eren F, Akkiprik M, Atug O, Sonmez O, Tahan G, Ozdemir F, et al. R72P polymorphism of TP53 in ulcerative colitis patients is associated with the incidence of colectomy, use of steroids and the presence of a positive family history. Pathol Oncol Res. 2010; **16**(4): 563-8.
28. Esteller M, Garcia A, Martinez-Palones JM, Xercavins J, Reventos J. Susceptibility to endometrial cancer: influence of allelism at p53, glutathione S-transferase (GSTM1 and GSTT1) and cytochrome P-450 (CYP1A1) loci. Br J Cancer. 1997; **75**(9): 1385-8.
29. Fernandez-Rubio A, Lopez-Cima MF, Gonzalez-Arriaga P, Garcia-Castro L, Pascual T, Marron MG, et al. The TP53 Arg72Pro polymorphism and lung cancer risk in a population of Northern Spain. Lung Cancer. 2008; **61**(3): 309-16.
30. Franekova M, Zubor P, Stanclova A, Dussan CA, Bohusova T, Galo S, et al. Association of p53 polymorphisms with breast cancer: a case-control study in Slovak population. Neoplasma. 2007; **54**(2): 155-61.

31. Garcia-Closas M, Kristensen V, Langerod A, Qi Y, Yeager M, Burdett L, et al. Common genetic variation in TP53 and its flanking genes, WDR79 and ATP1B2, and susceptibility to breast cancer. Int J Cancer. 2007; **121**(11): 2532-8.
32. Gaudet MM, Gammon MD, Bensen JT, Sagiv SK, Shantakumar S, Teitelbaum SL, et al. Genetic variation of TP53, polycyclic aromatic hydrocarbon-related exposures, and breast cancer risk among women on Long Island, New York. Breast Cancer Res Treat. 2008; **108**(1): 93-9.
33. Gemignani F, Moreno V, Landi S, Moullan N, Chabrier A, Gutierrez-Enriquez S, et al. A TP53 polymorphism is associated with increased risk of colorectal cancer and with reduced levels of TP53 mRNA. Oncogene. 2004; **23**(10): 1954-6.
34. Ghasemi N, Karimi-Zarchi M, Mortazavi-Zadeh MR, Atash-Afza A. Evaluation of the frequency of TP53 gene codon 72 polymorphisms in Iranian patients with endometrial cancer. Cancer Genet Cytogenet. 2010; **196**(2): 167-70.
35. Giuliani L, Jaxmar T, Casadio C, Gariglio M, Manna A, D'Antonio D, et al. Detection of oncogenic viruses SV40, BKV, JCV, HCMV, HPV and p53 codon 72 polymorphism in lung carcinoma. Lung Cancer. 2007; **57**(3): 273-81.
36. Gochhait S, Bukhari SI, Bairwa N, Vadhera S, Darvishi K, Raish M, et al. Implication of BRCA2 -26G>A 5' untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg>Pro polymorphism. Breast Cancer Res. 2007; **9**(5): R71.
37. Grunhage F, Jungck M, Lamberti C, Berg C, Becker U, Schulte-Witte H, et al. Association of familial colorectal cancer with variants in the E-cadherin (CDH1) and cyclin D1 (CCND1) genes. Int J Colorectal Dis. 2008; **23**(2): 147-54.
38. Hamajima N, Matsuo K, Suzuki T, Nakamura T, Matsuura A, Hatooka S, et al. No associations of p73 G4C14-to-A4T14 at exon 2 and p53 Arg72Pro polymorphisms with the risk of digestive tract cancers in Japanese. Cancer Lett. 2002; **181**(1): 81-5.
39. Helland A, Langerod A, Johnsen H, Olsen AO, Skovlund E, Borresen-Dale AL. p53 polymorphism and risk of cervical cancer. Nature. 1998; **396**(6711): 530-1; author reply 2.
40. Henriquez-Hernandez LA, Murias-Rosales A, Hernandez Gonzalez A, Cabrera De Leon A, Diaz-Chico BN, Mori De Santiago M, et al. Gene polymorphisms in TYMS, MTHFR, p53 and MDR1 as risk factors for breast cancer: a case-control study. Oncol Rep. 2009; **22**(6): 1425-33.
41. Hiraki A, Matsuo K, Hamajima N, Ito H, Hatooka S, Suyama M, et al. Different risk relations with smoking for non-small-cell lung cancer: comparison of TP53 and TP73 genotypes. Asian Pac J Cancer Prev. 2003; **4**(2): 107-12.
42. Honma HN, De Capitani EM, Perroud MW, Jr., Barbeiro AS, Toro IF, Costa DB, et al. Influence of p53 codon 72 exon 4, GSTM1, GSTT1 and GSTP1*B polymorphisms in lung cancer risk in a Brazilian population. Lung Cancer. 2008; **61**(2): 152-62.
43. Hrstka R, Beranek M, Klocova K, Nenutil R, Vojtesek B. Intronic polymorphisms in TP53 indicate lymph node metastasis in breast cancer. Oncol Rep. 2009; **22**(5): 1205-11.
44. Huang XE, Hamajima N, Katsuda N, Matsuo K, Hirose K, Mizutani M, et al. Association of p53 codon Arg72Pro and p73 G4C14-to-A4T14 at exon 2 genetic polymorphisms with the risk of Japanese breast cancer. Breast Cancer. 2003; **10**(4): 307-11.

45. Hung RJ, Boffetta P, Canzian F, Moullan N, Szeszenia-Dabrowska N, Zaridze D, et al. Sequence variants in cell cycle control pathway, X-ray exposure, and lung cancer risk: a multicenter case-control study in Central Europe. Cancer Res. 2006; **66**(16): 8280-6.
46. Ihsan R, Devi TR, Yadav DS, Mishra AK, Sharma J, Zomawia E, et al. Investigation on the Role of p53 Codon 72 Polymorphism and Interactions with Tobacco, Betel Quid, and Alcohol in Susceptibility to Cancers in a High-Risk Population from North East India. DNA Cell Biol. 2011; **30**(3): 163-71.
47. Irarrazabal CE, Rojas C, Aracena R, Marquez C, Gil L. Chilean pilot study on the risk of lung cancer associated with codon 72 polymorphism in the gene of protein p53. Toxicol Lett. 2003; **144**(1): 69-76.
48. Jain N, Singh V, Hedau S, Kumar S, Daga MK, Dewan R, et al. Infection of human papillomavirus type 18 and p53 codon 72 polymorphism in lung cancer patients from India. Chest. 2005; **128**(6): 3999-4007.
49. Jin X, Wu X, Roth JA, Amos CI, King TM, Branch C, et al. Higher lung cancer risk for younger African-Americans with the Pro/Pro p53 genotype. Carcinogenesis. 1995; **16**(9): 2205-8.
50. Joshi AM, Budhathoki S, Ohnaka K, Mibu R, Tanaka M, Kakeji Y, et al. TP53 R72P and MDM2 SNP309 polymorphisms and colorectal cancer risk: the Fukuoka Colorectal Cancer Study. Jpn J Clin Oncol. 2011; **41**(2): 232-8.
51. Jung HY, Whang YM, Sung JS, Shin HD, Park BL, Kim JS, et al. Association study of TP53 polymorphisms with lung cancer in a Korean population. J Hum Genet. 2008; **53**(6): 508-14.
52. Kalemi TG, Lambropoulos AF, Gueorguiev M, Chrisafi S, Papazisis KT, Kotsis A. The association of p53 mutations and p53 codon 72, Her 2 codon 655 and MTHFR C677T polymorphisms with breast cancer in Northern Greece. Cancer Lett. 2005; **222**(1): 57-65.
53. Kara N, Karakus N, Ulusoy AN, Ozaslan C, Gungor B, Bagci H. P53 codon 72 and HER2 codon 655 polymorphisms in Turkish breast cancer patients. DNA Cell Biol. 2010; **29**(7): 387-92.
54. Katiyar S, Thelma BK, Murthy NS, Hedau S, Jain N, Gopalkrishna V, et al. Polymorphism of the p53 codon 72 Arg/Pro and the risk of HPV type 16/18-associated cervical and oral cancer in India. Mol Cell Biochem. 2003; **252**(1-2): 117-24.
55. Kaushal M, Mishra AK, Raju BS, Ihsan R, Chakraborty A, Sharma J, et al. Betel quid chewing as an environmental risk factor for breast cancer. Mutat Res. 2010; **703**(2): 143-8.
56. Kawajiri K, Nakachi K, Imai K, Watanabe J, Hayashi S. Germ line polymorphisms of p53 and CYP1A1 genes involved in human lung cancer. Carcinogenesis. 1993; **14**(6): 1085-9.
57. Kazemi M, Salehi Z, Chakosari RJ. TP53 codon 72 polymorphism and breast cancer in northern Iran. Oncol Res. 2009; **18**(1): 25-30.
58. Keshava C, Frye BL, Wolff MS, McCanlies EC, Weston A. Waf-1 (p21) and p53 polymorphisms in breast cancer. Cancer Epidemiol Biomarkers Prev. 2002; **11**(1): 127-30.
59. Khadang B, Fattahi MJ, Talei A, Dehaghani AS, Ghaderi A. Polymorphism of TP53 codon 72 showed no association with breast cancer in Iranian women. Cancer Genet Cytogenet. 2007; **173**(1): 38-42.

60. Khaliq S, Hameed A, Khaliq T, Ayub Q, Qamar R, Mohyuddin A, et al. P53 mutations, polymorphisms, and haplotypes in Pakistani ethnic groups and breast cancer patients. Genet Test. 2000; **4**(1): 23-9.
61. Kiyohara C, Horiuchi T, Miyake Y, Takayama K, Nakanishi Y. Cigarette smoking, TP53 Arg72Pro, TP53BP1 Asp353Glu and the risk of lung cancer in a Japanese population. Oncol Rep. 2010; **23**(5): 1361-8.
62. Koushik A, Tranah GJ, Ma J, Stampfer MJ, Sesso HD, Fuchs CS, et al. p53 Arg72Pro polymorphism and risk of colorectal adenoma and cancer. Int J Cancer. 2006; **119**(8): 1863-8.
63. Kruger S, Bier A, Engel C, Mangold E, Pagenstecher C, von Knebel Doeberitz M, et al. The p53 codon 72 variation is associated with the age of onset of hereditary non-polyposis colorectal cancer (HNPCC). J Med Genet. 2005; **42**(10): 769-73.
64. Lang A, Palmeback Wegman P, Wingren S. The significance of MDM2 SNP309 and p53 Arg72Pro in young women with breast cancer. Oncol Rep. 2009; **22**(3): 575-9.
65. Li T, Lu ZM, Guo M, Wu QJ, Chen KN, Xing HP, et al. p53 codon 72 polymorphism (C/G) and the risk of human papillomavirus-associated carcinomas in China. Cancer. 2002; **95**(12): 2571-6.
66. Liu G, Miller DP, Zhou W, Thurston SW, Fan R, Xu LL, et al. Differential association of the codon 72 p53 and GSTM1 polymorphisms on histological subtype of non-small cell lung carcinoma. Cancer Res. 2001; **61**(24): 8718-22.
67. Lum SS, Chua HW, Li H, Li WF, Rao N, Wei J, et al. MDM2 SNP309 G allele increases risk but the T allele is associated with earlier onset age of sporadic breast cancers in the Chinese population. Carcinogenesis. 2008; **29**(4): 754-61.
68. Ma H, Hu Z, Zhai X, Wang S, Wang X, Qin J, et al. Joint effects of single nucleotide polymorphisms in P53BP1 and p53 on breast cancer risk in a Chinese population. Carcinogenesis. 2006; **27**(4): 766-71.
69. Mabrouk I, Baccouche S, El-Abed R, Mokdad-Gargouri R, Mosbah A, Said S, et al. No evidence of correlation between p53 codon 72 polymorphism and risk of bladder or breast carcinoma in Tunisian patients. Ann N Y Acad Sci. 2003; **1010**: 764-70.
70. Mahasneh AA, Abdel-Hafiz SS. Polymorphism of p53 gene in Jordanian population and possible associations with breast cancer and lung adenocarcinoma. Saudi Med J. 2004; **25**(11): 1568-73.
71. Matullo G, Dunning AM, Guarrera S, Baynes C, Polidoro S, Garte S, et al. DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. Carcinogenesis. 2006; **27**(5): 997-1007.
72. Mechanic LE, Bowman ED, Welsh JA, Khan MA, Hagiwara N, Enewold L, et al. Common genetic variation in TP53 is associated with lung cancer risk and prognosis in African Americans and somatic mutations in lung tumors. Cancer Epidemiol Biomarkers Prev. 2007; **16**(2): 214-22.
73. Menzel HJ, Sarmanova J, Soucek P, Berberich R, Grunewald K, Haun M, et al. Association of NQO1 polymorphism with spontaneous breast cancer in two independent populations. Br J Cancer. 2004; **90**(10): 1989-94.
74. Mojtahedi Z, Haghshenas MR, Hosseini SV, Fattahi MJ, Ghaderi A. p 53 codon 72 polymorphism in stomach and colorectal adenocarcinomas in Iranian patients. Indian J Cancer. 2010; **47**(1): 31-4.

75. Morari EC, Lima AB, Bufalo NE, Leite JL, Granja F, Ward LS. Role of glutathione-S-transferase and codon 72 of P53 genotypes in epithelial ovarian cancer patients. J Cancer Res Clin Oncol. 2006; **132**(8): 521-8.
76. Murata M, Tagawa M, Kimura M, Kimura H, Watanabe S, Saisho H. Analysis of a germ line polymorphism of the p53 gene in lung cancer patients; discrete results with smoking history. Carcinogenesis. 1996; **17**(2): 261-4.
77. Nadji SA, Mahmoodi M, Ziaee AA, Naghshvar F, Torabizadeh J, Yahyapour Y, et al. An increased lung cancer risk associated with codon 72 polymorphism in the TP53 gene and human papillomavirus infection in Mazandaran province, Iran. Lung Cancer. 2007; **56**(2): 145-51.
78. Niwa Y, Hirose K, Matsuo K, Tajima K, Ikoma Y, Nakanishi T, et al. Association of p73 G4C14-to-A4T14 polymorphism at exon 2 and p53 Arg72Pro polymorphism with the risk of endometrial cancer in Japanese subjects. Cancer Lett. 2005; **219**(2): 183-90.
79. Noma C, Miyoshi Y, Taguchi T, Tamaki Y, Noguchi S. Association of p53 genetic polymorphism (Arg72Pro) with estrogen receptor positive breast cancer risk in Japanese women. Cancer Lett. 2004; **210**(2): 197-203.
80. Nordgard SH, Alnaes GI, Hihn B, Lingjaerde OC, Liestol K, Tsalenko A, et al. Pathway based analysis of SNPs with relevance to 5-FU therapy: relation to intratumoral mRNA expression and survival. Int J Cancer. 2008; **123**(3): 577-85.
81. Ohayon T, Gershoni-Baruch R, Papa MZ, Distelman Menachem T, Eisenberg Barzilai S, Friedman E. The R72P P53 mutation is associated with familial breast cancer in Jewish women. Br J Cancer. 2005; **92**(6): 1144-8.
82. Olschwang S, Laurent-Puig P, Vassal A, Salmon RJ, Thomas G. Characterization of a frequent polymorphism in the coding sequence of the Tp53 gene in colonic cancer patients and a control population. Hum Genet. 1991; **86**(4): 369-70.
83. Papadakis ED, Soulitzis N, Spandidos DA. Association of p53 codon 72 polymorphism with advanced lung cancer: the Arg allele is preferentially retained in tumours arising in Arg/Pro germline heterozygotes. Br J Cancer. 2002; **87**(9): 1013-8.
84. Papadakis EN, Dokianakis DN, Spandidos DA. p53 codon 72 polymorphism as a risk factor in the development of breast cancer. Mol Cell Biol Res Commun. 2000; **3**(6): 389-92.
85. Park JK, Lee HJ, Kim JW, Park YH, Lee SS, Chang HI, et al. Differences in p53 gene polymorphisms between Korean schizophrenia and lung cancer patients. Schizophr Res. 2004; **67**(1): 71-4.
86. Pegoraro RJ, Moodley M, Rom L, Chetty R, Moodley J. P53 codon 72 polymorphism and BRCA 1 and 2 mutations in ovarian epithelial malignancies in black South Africans. Int J Gynecol Cancer. 2003; **13**(4): 444-9.
87. Peller S, Halperin R, Schneider D, Kopilova Y, Rotter V. Polymorphisms of the p53 gene in women with ovarian or endometrial carcinoma. Oncol Rep. 1999; **6**(1): 193-7.
88. Perez LO, Abba MC, Dulout FN, Golijow CD. Evaluation of p53 codon 72 polymorphism in adenocarcinomas of the colon and rectum in La Plata, Argentina. World J Gastroenterol. 2006; **12**(9): 1426-9.
89. Perfumo C, Bonelli L, Menichini P, Inga A, Gismondi V, Ciferri E, et al. Increased risk of colorectal adenomas in Italian subjects carrying the p53 PIN3 A2-Pro72 haplotype. Digestion. 2006; **74**(3-4): 228-35.

90. Piao JM, Kim HN, Song HR, Kweon SS, Choi JS, Yun WJ, et al. p53 codon 72 polymorphism and the risk of lung cancer in a Korean population. Lung Cancer. 2011.
91. Pierce LM, Sivaraman L, Chang W, Lum A, Donlon T, Seifried A, et al. Relationships of TP53 codon 72 and HRAS1 polymorphisms with lung cancer risk in an ethnically diverse population. Cancer Epidemiol Biomarkers Prev. 2000; **9**(11): 1199-204.
92. Polakova V, Pardini B, Naccarati A, Landi S, Slyskova J, Novotny J, et al. Genotype and haplotype analysis of cell cycle genes in sporadic colorectal cancer in the Czech Republic. Hum Mutat. 2009; **30**(4): 661-8.
93. Popanda O, Edler L, Waas P, Schattenberg T, Butkiewicz D, Muley T, et al. Elevated risk of squamous-cell carcinoma of the lung in heavy smokers carrying the variant alleles of the TP53 Arg72Pro and p21 Ser31Arg polymorphisms. Lung Cancer. 2007; **55**(1): 25-34.
94. Rajkumar T, Samson M, Rama R, Sridevi V, Mahji U, Swaminathan R, et al. TGFbeta1 (Leu10Pro), p53 (Arg72Pro) can predict for increased risk for breast cancer in south Indian women and TGFbeta1 Pro (Leu10Pro) allele predicts response to neo-adjuvant chemo-radiotherapy. Breast Cancer Res Treat. 2008; **112**(1): 81-7.
95. Sakiyama T, Kohno T, Mimaki S, Ohta T, Yanagitani N, Sobue T, et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. Int J Cancer. 2005; **114**(5): 730-7.
96. Sameer AS, Shah ZA, Syeed N, Banday MZ, Bashir SM, Bhat BA, et al. TP53 Pro47Ser and Arg72Pro polymorphisms and colorectal cancer predisposition in an ethnic Kashmiri population. Genet Mol Res. 2010; **9**(2): 651-60.
97. Samson M, Swaminathan R, Rama R, Sridevi V, Nancy KN, Rajkumar T. Role of GSTM1 (Null/Present), GSTP1 (Ile105Val) and P53 (Arg72Pro) genetic polymorphisms and the risk of breast cancer: a case control study from South India. Asian Pac J Cancer Prev. 2007; **8**(2): 253-7.
98. Santos AM, Sousa H, Pinto D, Portela C, Pereira D, Catarino R, et al. Linking TP53 codon 72 and P21 nt590 genotypes to the development of cervical and ovarian cancer. Eur J Cancer. 2006; **42**(7): 958-63.
99. Sayhan N, Yazici H, Budak M, Bitisik O, Dalay N. P53 codon 72 genotypes in colon cancer. Association with human papillomavirus infection. Res Commun Mol Pathol Pharmacol. 2001; **109**(1-2): 25-34.
100. Schildkraut JM, Goode EL, Clyde MA, Iversen ES, Moorman PG, Berchuck A, et al. Single nucleotide polymorphisms in the TP53 region and susceptibility to invasive epithelial ovarian cancer. Cancer Res. 2009; **69**(6): 2349-57.
101. Schmidt MK, Reincke S, Broeks A, Braaf LM, Hogervorst FB, Tollenaar RA, et al. Do MDM2 SNP309 and TP53 R72P interact in breast cancer susceptibility? A large pooled series from the breast cancer association consortium. Cancer Res. 2007; **67**(19): 9584-90.
102. Schneider-Stock R, Boltze C, Peters B, Szibor R, Landt O, Meyer F, et al. Selective loss of codon 72 proline p53 and frequent mutational inactivation of the retained arginine allele in colorectal cancer. Neoplasia. 2004; **6**(5): 529-35.
103. Schneider-Stock R, Mawrin C, Motsch C, Boltze C, Peters B, Hartig R, et al. Retention of the arginine allele in codon 72 of the p53 gene correlates with poor apoptosis in head and neck cancer. Am J Pathol. 2004; **164**(4): 1233-41.

104. Sergentanis TN, Economopoulos KP. Eligible and not eligible studies in the recent meta-analysis about p53 polymorphism and breast cancer risk. Breast Cancer Res Treat. 2010; **120**(1): 261-2.
105. Siddique MM, Balram C, Fiszer-Maliszewska L, Aggarwal A, Tan A, Tan P, et al. Evidence for selective expression of the p53 codon 72 polymorphs: implications in cancer development. Cancer Epidemiol Biomarkers Prev. 2005; **14**(9): 2245-52.
106. Sigurdson AJ, Bhatti P, Doody MM, Hauptmann M, Bowen L, Simon SL, et al. Polymorphisms in apoptosis- and proliferation-related genes, ionizing radiation exposure, and risk of breast cancer among U.S. Radiologic Technologists. Cancer Epidemiol Biomarkers Prev. 2007; **16**(10): 2000-7.
107. Singh V, Rastogi N, Mathur N, Singh K, Singh MP. Association of polymorphism in MDM-2 and p53 genes with breast cancer risk in Indian women. Ann Epidemiol. 2008; **18**(1): 48-57.
108. Sjalander A, Birgander R, Athlin L, Stenling R, Rutegard J, Beckman L, et al. P53 germ line haplotypes associated with increased risk for colorectal cancer. Carcinogenesis. 1995; **16**(7): 1461-4.
109. Sjalander A, Birgander R, Hallmans G, Cajander S, Lenner P, Athlin L, et al. p53 polymorphisms and haplotypes in breast cancer. Carcinogenesis. 1996; **17**(6): 1313-6.
110. Smeds J, Nava M, Kumar R, Hemminki K. A novel polymorphism (-88 C>A) in the 5' UTR of the p53R2 gene. Hum Mutat. 2001; **17**(1): 82.
111. Sobti RC, Kaur P, Kaur S, Janmeja AK, Jindal SK, Kishan J, et al. Impact of interaction of polymorphic forms of p53 codon 72 and N-acetylation gene (NAT2) on the risk of lung cancer in the North Indian population. DNA Cell Biol. 2009; **28**(9): 443-9.
112. Song F, Zheng H, Liu B, Wei S, Dai H, Zhang L, et al. An miR-502-binding site single-nucleotide polymorphism in the 3'-untranslated region of the SET8 gene is associated with early age of breast cancer onset. Clin Cancer Res. 2009; **15**(19): 6292-300.
113. Sotamaa K, Liyanarachchi S, Mecklin JP, Jarvinen H, Aaltonen LA, Peltomaki P, et al. p53 codon 72 and MDM2 SNP309 polymorphisms and age of colorectal cancer onset in Lynch syndrome. Clin Cancer Res. 2005; **11**(19 Pt 1): 6840-4.
114. Sprague BL, Trentham-Dietz A, Garcia-Closas M, Newcomb PA, Titus-Ernstoff L, Hampton JM, et al. Genetic variation in TP53 and risk of breast cancer in a population-based case control study. Carcinogenesis. 2007; **28**(8): 1680-6.
115. Sreeja L, Syamala V, Raveendran PB, Santhi S, Madhavan J, Ankathil R. p53 Arg72Pro polymorphism predicts survival outcome in lung cancer patients in Indian population. Cancer Invest. 2008; **26**(1): 41-6.
116. Suspitsin EN, Buslov KG, Grigoriev MY, Ishutkina JG, Ulibina JM, Gorodinskaya VM, et al. Evidence against involvement of p53 polymorphism in breast cancer predisposition. Int J Cancer. 2003; **103**(3): 431-3.
117. Szymanowska A, Jassem E, Dziadziuszko R, Borg A, Limon J, Kobierska-Gulida G, et al. Increased risk of non-small cell lung cancer and frequency of somatic TP53 gene mutations in Pro72 carriers of TP53 Arg72Pro polymorphism. Lung Cancer. 2006; **52**(1): 9-14.
118. Tan XL, Nieters A, Hoffmeister M, Beckmann L, Brenner H, Chang-Claude J. Genetic polymorphisms in TP53, nonsteroidal anti-inflammatory drugs and the risk of

colorectal cancer: evidence for gene-environment interaction? Pharmacogenet Genomics. 2007; **17**(8): 639-45.
119. Tarasov VA, Aslanyan MM, Tsyrendorzhiyeva ES, Litvinov SS, Gar'kavtseva RF, Altukhov YP. Genetically determined subdivision of human populations with respect to the risk of breast cancer in women. Dokl Biol Sci. 2006; **406**: 66-9.
120. To-Figueras J, Gene M, Gomez-Catalan J, Galan C, Firvida J, Fuentes M, et al. Glutathione-S-Transferase M1 and codon 72 p53 polymorphisms in a northwestern Mediterranean population and their relation to lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev. 1996; **5**(5): 337-42.
121. Tommiska J, Eerola H, Heinonen M, Salonen L, Kaare M, Tallila J, et al. Breast cancer patients with p53 Pro72 homozygous genotype have a poorer survival. Clin Cancer Res. 2005; **11**(14): 5098-103.
122. Trifa F, Karray-Chouayekh S, Mabrouk I, Baccouche S, Khabir A, Sellami-Boudawara T, et al. Haplotype analysis of p53 polymorphisms: Arg72Pro, Ins16bp and G13964C in Tunisian patients with familial or sporadic breast cancer. Cancer Epidemiol. 2010; **34**(2): 184-8.
123. Ueda M, Terai Y, Kanda K, Kanemura M, Takehara M, Yamaguchi H, et al. Germline polymorphism of p53 codon 72 in gynecological cancer. Gynecol Oncol. 2006; **100**(1): 173-8.
124. Wang NM, Tsai CH, Yeh KT, Chen SJ, Chang JG. P53 codon 72Arg polymorphism is not a risk factor for carcinogenesis in the chinese. Int J Mol Med. 1999; **4**(3): 249-52.
125. Wang W, Spitz MR, Yang H, Lu C, Stewart DJ, Wu X. Genetic variants in cell cycle control pathway confer susceptibility to lung cancer. Clin Cancer Res. 2007; **13**(19): 5974-81.
126. Wang YC, Chen CY, Chen SK, Chang YY, Lin P. p53 codon 72 polymorphism in Taiwanese lung cancer patients: association with lung cancer susceptibility and prognosis. Clin Cancer Res. 1999; **5**(1): 129-34.
127. Wang-Gohrke S, Rebbeck TR, Besenfelder W, Kreienberg R, Runnebaum IB. p53 germline polymorphisms are associated with an increased risk for breast cancer in German women. Anticancer Res. 1998; **18**(3B): 2095-9.
128. Wasik M, Gorska E, Popko K, Pawelec K, Matysiak M, Demkow U. The Gln223Arg polymorphism of the leptin receptor gene and peripheral blood/bone marrow leptin level in leukemic children. J Physiol Pharmacol. 2006; **57 Suppl 4**: 375-83.
129. Webb EL, Rudd MF, Sellick GS, El Galta R, Bethke L, Wood W, et al. Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. Hum Mol Genet. 2006; **15**(21): 3263-71.
130. Weston A, Pan CF, Ksieski HB, Wallenstein S, Berkowitz GS, Tartter PI, et al. p53 haplotype determination in breast cancer. Cancer Epidemiol Biomarkers Prev. 1997; **6**(2): 105-12.
131. Weston A, Perrin LS, Forrester K, Hoover RN, Trump BF, Harris CC, et al. Allelic frequency of a p53 polymorphism in human lung cancer. Cancer Epidemiol Biomarkers Prev. 1992; **1**(6): 481-3.

132. Zhang X, Miao X, Guo Y, Tan W, Zhou Y, Sun T, et al. Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. Hum Mutat. 2006; **27**(1): 110-7.
133. Zhu ZZ, Wang AZ, Jia HR, Jin XX, He XL, Hou LF, et al. Association of the TP53 codon 72 polymorphism with colorectal cancer in a Chinese population. Jpn J Clin Oncol. 2007; **37**(5): 385-90.
134. Zubor P, Stanclova A, Kajo K, Hatok J, Klobusiakova D, Visnovsky J, et al. The p53 codon 72 exon 4 BstUI polymorphism and endometrial cancer in Caucasian women. Oncology. 2009; **76**(3): 173-83.

Appendix Document 3: Model specification for maximum likelihood and Bayesian analyses.

2-level model (cancer as a covariate)

We wanted to examine the effect of using DNA isolated from tumor tissue vs. normal tissue on the association of the Arg72Pro polymorphism with five common epithelial cancers (lung, breast, colorectal, ovarian, endometrial). Because of loss-of-heterozygosity (LOH), associations in tumor tissue are spurious (biased). We assumed that LOH has the same effect (biases the association by a similar amount) across the three cancer topics; however, we allowed the genetic effect to be different for each cancer subtype.

We specified a two level model: the first level was the patient. The second level is the study. At the second level we have parameters that are common across all five cancer topics. Subscript i = [1,2,…,N] denotes the study. N is the number of published studies.

Level one (within studies)

For the cases and controls in the i-th study we assume that the number of Pro-encoding alleles, r, follows a binomial distribution with probability p (i.e., the true frequency of the minor allele) out of a sample of size n (the total number of alleles in each group):

$$r_i^{cas} \sim Bin(p_i^{cas}, n_i^{case}) \text{ and } r_i^{con} \sim Bin(p_i^{con}, n_i^{con})$$

We assume that the log-transformed odds ratios, d_i, of each study

$$d_i = \log\frac{p_i^{cas}}{1-p_i^{cas}} - \log\frac{p_i^{con}}{1-p_i^{con}}$$

are random effects that vary by function of a study specific genetic effect, d_i^*, and its modification by explanatory variables x_i (an indicator of the use of tumor or normal tissue for genotyping), w_i (an indicator of whether the study investigated lung cancer), z_i (an indicator of whether the study investigated colorectal cancer), v_i (an indicator of whether the study investigated ovarian cancer) and g_i (an indicator of whether the study investigated endometrial cancer); breast cancer is not encoded by an indicator variable because it serves as the baseline cancer type in this analysis

$$d_i = d_i^* + b_{tissue} \cdot x_i + b_{lung,i} \cdot w_i + b_{colorectal,i} \cdot z_i + b_{ovarian,i} \cdot v_i + b_{endometrial,i} \cdot g_i$$

Here b_{tissue}, b_{lung}, $b_{colorectal}$, $b_{ovarian}$ and $b_{endmetrial}$ are the regression coefficients for the indicator variables for use of tumor tissue, lung cancer, colorectal cancer, ovarian cancer and endometrial cancer, respectively. d^* is technically the log odds ratio for breast cancer studies.

Level two (study level)

We assume that $d^* \sim N(\bar{\delta}, \tau^2)$, $b_{lung} \sim N(\overline{\beta_{lung}}, \tau^2)$, $b_{colorectal} \sim N(\overline{\beta_{colorectal}}, \tau^2)$, $b_{ovarian} \sim N(\overline{\beta_{ovarian}}, \tau^2)$, $b_{endometrial} \sim N(\overline{\beta_{endometrial}}, \tau^2)$ and $\log \frac{p_i^{con}}{1-p_i^{con}} \sim N(\bar{\mu}, \upsilon^2)$, where $\bar{\delta}$ is the log-transformed "overall" summary effect in breast cancer and $\bar{\delta} + \overline{\beta_{lung}}$, $\bar{\delta} + \overline{\beta_{colorectal}}$, $\bar{\delta} + \overline{\beta_{ovarian}}$, $\bar{\delta} + \overline{\beta_{endmetrial}}$ are the summary log-transformed effect sizes for lung, colorectal, ovarian and endometrial cancer, respectively. $\bar{\mu}$ is the summary logit-transformed frequency of the minor allele in controls (a nuisance parameter).

In this analysis the bias effect was treated as a fixed effect across studies with prior distribution $b_{tissue} \sim N(0,10^6)$. Analyses where the bias effect was treated as a random effect across studies produced similar results (not shown).

Non-informative priors for the above formulae: $\bar{\delta} \sim N(0,10^6)$, $\overline{\beta_{lung}} \sim N(0,10^6)$, $\overline{\beta_{colorectal}} \sim N(0,10^6)$, $\overline{\beta_{ovarian}} \sim N(0,10^6)$, $\overline{\beta_{endometrial}} \sim N(0,10^6)$, $\bar{\mu} \sim N(0,10^6)$, $\tau \sim U(0,10)$ and $\upsilon \sim U(0,10)$.

3-level model (cancer topics as an additional level)

We also specified a three level model with the first level within study, second level within cancer topic study and third level across topics. Again, subscript $j = [1, ..., 5]$ denotes the cancer topic, and subscript $i = [1, 2, ..., N_j]$ denotes the study. N_j is the number of published studies for the *i*-th cancer.

Level 1 (within studies)

For the cases and controls in the i-th study nested within the j-th cancer topic, we assume that the number of C alleles, *r*, follows a binomial distribution with probability *p* (i.e., the true frequency of the minor allele) out of a sample of size *n* (the total number of alleles in each group):

$$r_{i(j)}^{cas} \sim Bin(p_{i(j)}^{cas}, n_{i(j)}^{case}) \text{ and } r_{i(j)}^{con} \sim Bin(p_{i(j)}^{con}, n_{i(j)}^{con})$$

The log-transformed odds ratios, $d_{i(j)}$, of each study nested within a cancer topic

$$d_{i(j)} = \log \frac{p_{i(j)}^{cas}}{1 - p_{i(j)}^{cas}} - \log \frac{p_{i(j)}^{con}}{1 - p_{i(j)}^{con}}$$

are random effects with cancer-specific means that vary by whether or not the study employed tumor or normal tissue for genotyping

$$d_{i(j)} \sim N\left(\delta_j + b_{tissue,j} \cdot x_{i(j)}, \tau_j^2\right).$$

We also assume that a distribution for the random effects of the cancer specific summary logit-transformed frequency of the minor allele (a nuisance parameter)

$$\log \frac{p_{i(j)}^{con}}{1 - p_{i(j)}^{con}} \sim N(\mu_j, \upsilon_j^2).$$

Non-informative priors for the variances are $\tau_j \sim U(0,10)$ and $\upsilon_j \sim U(0,10)$.

Level 2 (within cancer topics)

We assume that the cancer-level bias is a random effect $b_{tissue,j} \sim N(\overline{\beta_{tissue}}, \tau_{tissue}^2)$, where $\overline{\beta_{tissue}}$ is the "common" bias effect incurred by the LOH phenomenon across cancer topics. Similarly, at the cancer-level, $\delta_j \sim N(\bar{\delta}, \sigma^2)$ and $\mu_j \sim N(0, 10^6)$, where $\bar{\delta}$ is the "common" true unbiased effect (log odds ratio) across all cancers.

Priors for the variances in the above formulae: $\sigma \sim U(0,10)$, and $\tau_{tissue} \sim U(0,10)$.

Level 3 (common parameters across topics)

The prior for he common bias effect incurred by the LOH phenomenon across all cancers is $\overline{\beta_{tissue}} \sim \mathcal{N}(0,10^6)$ and for the common "true" unbiased genetic effect is $\overline{\delta} \sim N(0,10^6)$.

Appendix Tables

Appendix Table 1

Meta-analysis results for breast cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Characteristic		Studies (cases, controls)	Heterogeneity (pQ; I^2)	OR (95% CI); p-value
All studies		60 (27728, 31073)	<0.001; 56%	0.98 (0.94-1.03); 0.465
Ethnicity	Whites	41 (23490, 25646)	<0.001; 52%	1.02 (0.97-1.07); 0.527
	East Asians	7 (2859, 2880)	0.055; 51%	1.04 (0.92-1.16); 0.525
Control selection	Disease controls	5 (1485, 1593)	0.224; 30%	1.13 (0.99-1.28); 0.077
	Healthy controls	55 (26243, 29480)	<0.001; 57%	0.97 (0.92-1.02); 0.219
Matching	No/NR	31 (16691, 17314)	0.075; 28%	0.98 (0.93-1.03); 0.399
	Yes	29 (11037, 13759)	<0.001; 69%	0.98 (0.90-1.06); 0.603
Genotyping QC	No/NR	38 (8643, 11923)	<0.001; 57%	0.95 (0.87-1.02); 0.175
	Yes	22 (19085, 19150)	0.001; 55%	1.01 (0.96-1.07); 0.670
Blinding	No	54 (23323, 25811)	<0.001; 58%	0.98 (0.92-1.03); 0.368
	Yes	6 (4405, 5262)	0.199; 32%	1.01 (0.94-1.09); 0.735
Genotyping method	Non-RFLP	30 (20795, 22849)	<0.001; 59%	0.97 (0.92-1.03); 0.314
	RFLP	30 (6933, 8224)	<0.001; 55%	1.00 (0.92-1.09); 0.968
HWE	Compliant	48 (26076, 28120)	<0.001; 53%	1.01 (0.96-1.05); 0.793
	In violation	12 (1652, 2953)	0.003; 61%	0.84 (0.68-1.00); 0.033

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; p_Q = p-value from Cochran's Q statistic; QC = quality control; RFLP = restriction fragment length polymorphism.

Appendix Table 2

Meta-analysis results for lung cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Characteristic		Studies (cases, controls)	Heterogeneity (pQ; I^2)	OR (95% CI); p-value
All studies		39 (16522, 16235)	<0.001; 54%	1.09 (1.03-1.15); 0.003
Ethnicity	Whites	15 (7121, 8596)	<0.001; 68%	1.04 (0.93-1.14); 0.521
	East Asians	12 (8028, 5877)	0.067; 41%	1.13 (1.05-1.21); 0.002
Control selection	Disease controls	18 (7209, 7472)	0.230; 19%	1.15 (1.09-1.21); 0.000
	Healthy controls	21 (9313, 8763)	<0.001; 67%	1.03 (0.93-1.12); 0.579
Matching	No/NR	18 (8951, 6965)	0.008; 50%	1.10 (1.02-1.18); 0.018
	Yes	21 (7571, 9270)	0.001; 55%	1.08 (1.00-1.17); 0.063
Genotyping QC	No/NR	22 (7488, 5545)	<0.001; 59%	1.07 (0.97-1.17); 0.178
	Yes	17 (9034, 10690)	0.018; 47%	1.10 (1.03-1.17); 0.006
Blinding	No	30 (12125, 11457)	<0.001; 55%	1.09 (1.02-1.16); 0.020
	Yes	9 (4397, 4778)	0.037; 51%	1.09 (0.99-1.19); 0.098
Genotyping method	Non-RFLP	16 (10801, 9502)	<0.001; 65%	1.09 (1.00-1.18); 0.064
	RFLP	23 (5721, 6733)	0.011; 45%	1.09 (1.01-1.18); 0.032
HWE	Compliant	34 (16034, 15519)	0.003; 45%	1.10 (1.04-1.15); 0.001
	In violation	5 (488, 716)	<0.001; 82%	0.95 (0.52-1.38); 0.810

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; p_Q = p-value from Cochran's Q statistic; QC = quality control; RFLP = restriction fragment length polymorphism.

Appendix Table 3

Meta-analysis results for colorectal cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Characteristic		Studies (cases, controls)	Heterogeneity (pQ; I^2)	OR (95% CI); p-value
All studies		20 (6951, 9275)	<0.001; 75%	1.09 (0.98-1.20); 0.136
Ethnicity	Whites	12 (5273, 6446)	0.358; 9%	0.97 (0.90-1.04); 0.351
	East Asians	7 (1592, 2669)	0.005; 67%	1.18 (1.00-1.35); 0.074
Control selection	Disease controls	7 (1537, 1935)	0.218; 28%	1.05 (0.91-1.18); 0.501
	Healthy controls	13 (5414, 7340)	<0.001; 83%	1.11 (0.96-1.26); 0.183
Matching	No/NR	13 (3962, 5272)	0.156; 29%	0.98 (0.88-1.08); 0.721
	Yes	7 (2989, 4003)	<0.001; 86%	1.25 (1.05-1.46); 0.033
Genotyping QC	No/NR	10 (1620, 2527)	0.002; 65%	1.07 (0.89-1.24); 0.456
	Yes	10 (5331, 6748)	<0.001; 81%	1.11 (0.95-1.26); 0.196
Blinding	No	17 (5886, 7515)	<0.001; 77%	1.10 (0.96-1.23); 0.171
	Yes	3 (1065, 1760)	0.033; 71%	1.06 (0.82-1.30); 0.633
Genotyping method	Non-RFLP	10 (4559, 5788)	0.266; 19%	0.95 (0.87-1.04); 0.252
	RFLP	10 (2392, 3487)	<0.001; 78%	1.25 (1.07-1.44); 0.016
HWE	Compliant	19 (6865, 9115)	<0.001; 62%	1.05 (0.95-1.14); 0.344
	In violation	1 (86, 160)	NA	3.07 (2.68-3.45); 0.000

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; NA = not applicable; OR = odds ratio; p_Q = p-value from Cochran's Q statistic; QC = quality control; RFLP = restriction fragment length polymorphism.

Appendix Table 4

Meta-analysis results for ovarian cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Characteristic		Studies (cases, controls)	Heterogeneity (pQ; I^2)	OR (95% CI); p-value
All studies		14 (1892, 5146)	0.534; 0%	1.10 (1.01-1.19); 0.031
Ethnicity	Whites	12 (1779, 4711)	0.567; 0%	1.11 (1.02-1.21); 0.019
	East Asians	1 (68, 95)	NA	1.15 (0.69-1.60); 0.558
Control selection	Disease controls	1 (45, 340)	NA	0.78 (0.32-1.24); 0.281
	Healthy controls	13 (1847, 4806)	0.650; 0%	1.12 (1.03-1.21); 0.016
Matching	No/NR	9 (1091, 2934)	0.674; 0%	1.12 (1.01-1.24); 0.050
	Yes	5 (801, 2212)	0.208; 32%	1.07 (0.90-1.23); 0.437
Genotyping QC	No/NR	6 (484, 910)	0.385; 5%	1.19 (0.99-1.39); 0.086
	Yes	8 (1408, 4236)	0.560; 0%	1.08 (0.98-1.18); 0.131
Blinding	No	13 (1700, 4691)	0.476; 0%	1.09 (1.00-1.18); 0.063
	Yes	1 (192, 455)	NA	1.18 (0.91-1.45); 0.224
Genotyping method	Non-RFLP	12 (1755, 4829)	0.463; 0%	1.09 (1.00-1.18); 0.071
	RFLP	2 (137, 317)	0.575; 0%	1.26 (0.96-1.57); 0.135
HWE	Compliant	11 (1512, 4071)	0.574; 0%	1.10 (1.00-1.20); 0.065
	In violation	3 (380, 1075)	0.190; 40%	1.12 (0.87-1.38); 0.372

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; NA = not applicable; OR = odds ratio; p_Q = p-value from Cochran's Q statistic; QC = quality control; RFLP = restriction fragment length polymorphism.

Appendix Table 5

Meta-analysis results for endometrial cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

Characteristic		Studies (cases, controls)	Heterogeneity (pQ; I^2)	OR (95% CI); p-value
All studies		6 (590, 1202)	0.265; 22%	1.10 (0.91-1.29); 0.338
Ethnicity	Whites	4 (368, 665)	0.739; 0%	1.22 (1.01-1.43); 0.068
	East Asians	2 (222, 537)	0.080; 67%	0.91 (0.46-1.36); 0.682
Control selection	Disease controls	2 (305, 732)	0.803; 0%	1.15 (0.94-1.36); 0.193
	Healthy controls	4 (285, 470)	0.114; 50%	1.08 (0.69-1.47); 0.697
Matching	No/NR	4 (278, 582)	0.119; 49%	1.07 (0.69-1.44); 0.735
	Yes	2 (312, 620)	0.912; 0%	1.17 (0.94-1.39); 0.175
Genotyping QC	No/NR	4 (278, 582)	0.119; 49%	1.07 (0.69-1.44); 0.735
	Yes	2 (312, 620)	0.912; 0%	1.17 (0.94-1.39); 0.175
Blinding	No	6 (590, 1202)	0.265; 22%	1.10 (0.91-1.29); 0.338
	Yes	none	NA	NA
Genotyping method	Non-RFLP	4 (361, 777)	0.700; 0%	1.19 (1.00-1.39); 0.080
	RFLP	2 (229, 425)	0.078; 68%	0.92 (0.44-1.39); 0.716
HWE	Compliant	5 (482, 1107)	0.834; 0%	1.18 (1.01-1.36); 0.055
	In violation	1 (108, 95)	NA	0.71 (0.29-1.13); 0.104

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; NA = not applicable; OR = odds ratio; p_Q = p-value from Cochran's Q statistic; QC = quality control; RFLP = restriction fragment length polymorphism.

Appendix Table 6

Meta-regression results for breast, lung and colorectal cancer, excluding studies using tumor tissue as the source of genotyping material for cases.

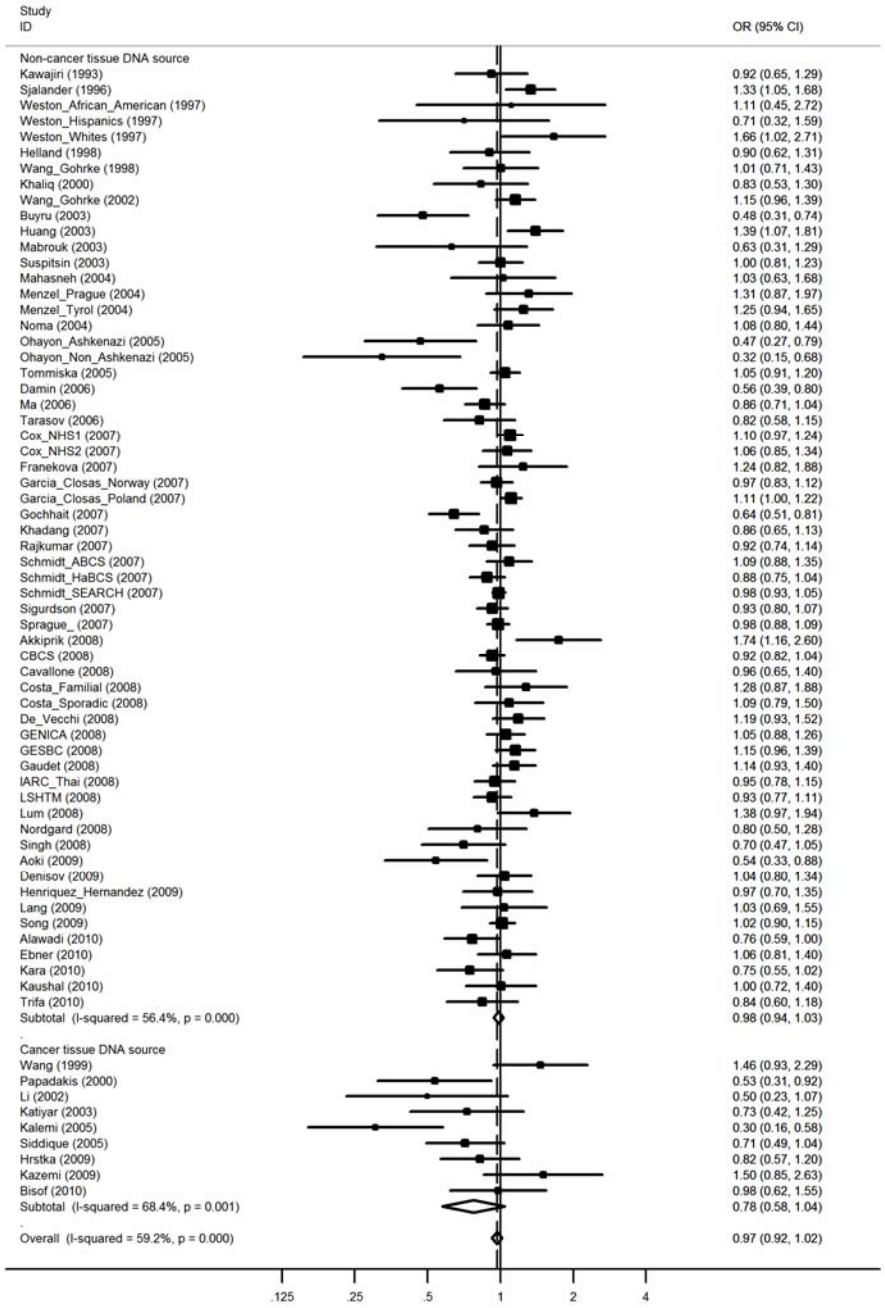
Cancer	*Contrast*	*rOR (95% CI)*	*p-value*
Breast cancer	East Asians versus Whites	1.02 (0.88-1.17)	0.821
	Disease versus healthy controls	0.85 (0.70-1.04)	0.119
	Matching versus no matching	1.01 (0.90-1.13)	0.889
	Genotyping QC versus no/NR	1.06 (0.95-1.19)	0.304
	Blinding verso no/NR	1.04 (0.88-1.22)	0.651
	RFLP versus non-RFLP method	1.04 (0.93-1.17)	0.512
	Violations versus compliance of HWE	**0.85 (0.73-0.99)**	**0.042**
	Year of publication (continuous)	0.99 (0.98-1.01)	0.337
Lung cancer	East Asians versus Whites	1.07 (0.91-1.25)	0.415
	Disease versus healthy controls	0.90 (0.79-1.02)	0.108
	Matching versus no matching	0.99 (0.86-1.12)	0.822
	Genotyping QC versus no/NR	1.02 (0.89-1.16)	0.807
	Blinding verso no/NR	1.00 (0.86-1.16)	0.977
	RFLP versus non-RFLP method	1.00 (0.88-1.15)	0.955
	Violations versus compliance of HWE	0.97 (0.76-1.24)	0.818
	Year of publication (continuous)	1.01 (0.99-1.02)	0.303
Colorectal cancer	East Asians versus Whites	**1.21 (1.03-1.44)**	**0.023**
	Disease versus healthy controls	1.05 (0.80-1.40)	0.711
	Matching versus no matching	1.25 (0.98-1.61)	0.077
	Genotyping QC versus no/NR	1.04 (0.80-1.36)	0.754
	Blinding verso no/NR	0.97 (0.68-1.39)	0.869
	RFLP versus non-RFLP method	**1.31 (1.04-1.65)**	**0.023**
	Violations versus compliance of HWE	**2.93 (1.79-4.80)**	**<0.001**
	Year of publication (continuous)	1.01 (0.99-1.04)	0.215
Ovarian cancer	East Asians versus Whites	1.03 (0.65-1.63)	0.908
	Disease versus healthy controls	1.44 (0.90-2.29)	0.129
	Matching versus no matching	0.95 (0.80-1.14)	0.607
	Genotyping QC versus no/NR	0.90 (0.73-1.12)	0.363
	Blinding verso no/NR	1.08 (0.81-1.43)	0.591
	RFLP versus non-RFLP method	1.16 (0.84-1.59)	0.361
	Violations versus compliance of HWE	1.02 (0.82-1.26)	0.864
	Year of publication (continuous)	1.00 (0.96-1.03)	0.814
Endometrial cancer	East Asians versus Whites	0.76 (0.52-1.12)	0.164
	Disease versus healthy controls	0.91 (0.59-1.42)	0.678
	Matching versus no matching	1.12 (0.73-1.73)	0.593
	Genotyping QC versus no/NR	1.12 (0.73-1.73)	0.593
	Blinding verso no/NR	1.10 (0.91-1.32)	0.321
	RFLP versus non-RFLP method	0.77 (0.53-1.13)	0.189
	Violations versus compliance of HWE	**0.60 (0.38-0.94)**	**0.026**
	Year of publication (continuous)	1.03 (0.94-1.12)	0.539

CI = confidence interval; HWE = Hardy-Weinberg equilibrium; NR = not reported; RFLP = restriction fragment length polymorphism; QC = quality control; rOR = relative odds ratio. Significant results are shown in bold type.

Appendix Figures

Appendix Figure 1

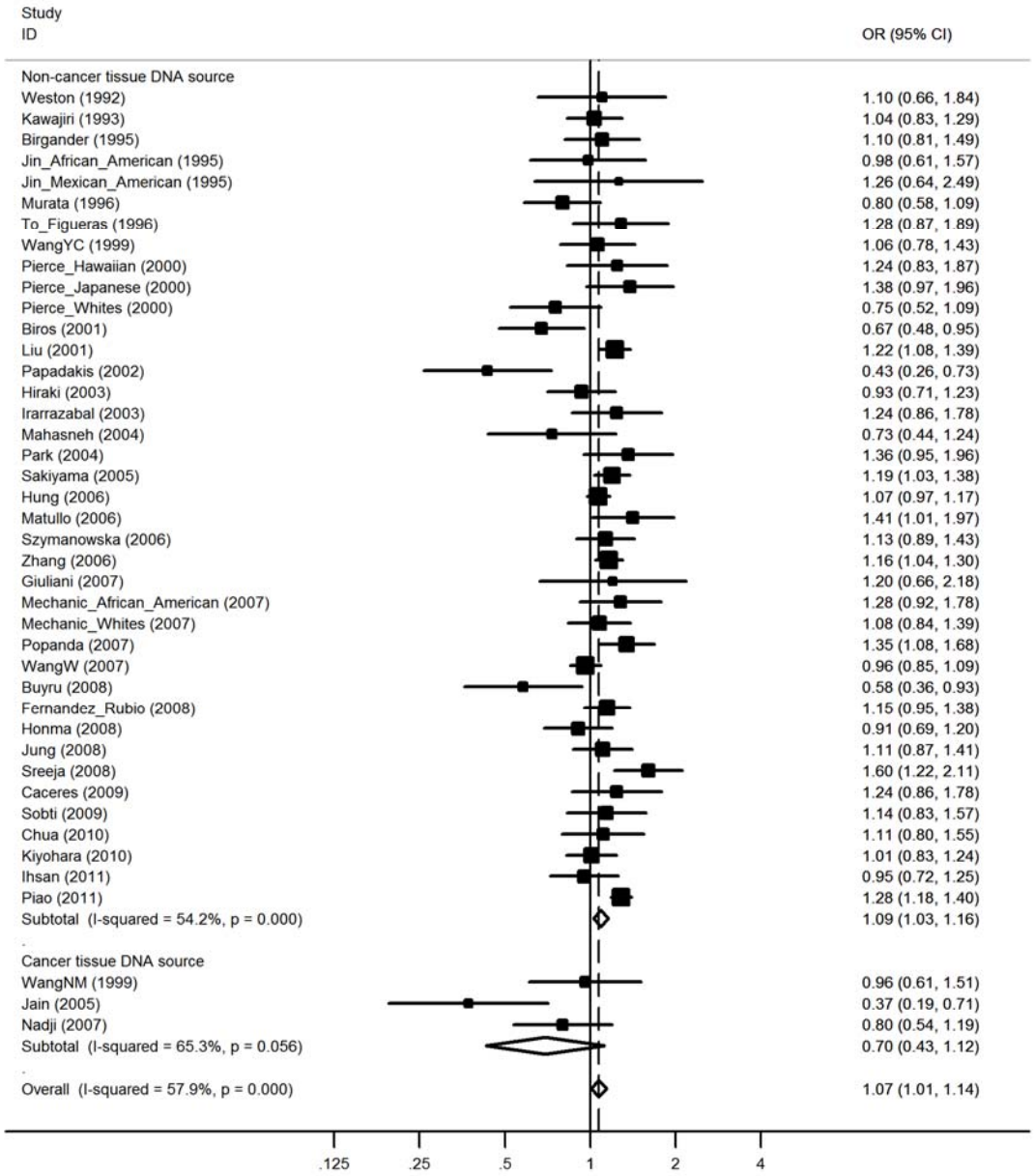
Forest plot for studies in breast cancer. Studies have been stratified by the source of DNA for genotyping cases.



Each study is shown by the point estimate of the odds ratio (square proportional to the weight of each study) and 95% confidence interval for the odds ratio (extending lines); the summary odds ratio and 95% confidence intervals by random effects calculations is depicted as a diamond. Values higher than 1 indicate that the Pro-encoding allele is associated with increased cancer risk. Studies are listed by year of publication and then alphabetically (by first author name).

Appendix Figure 2

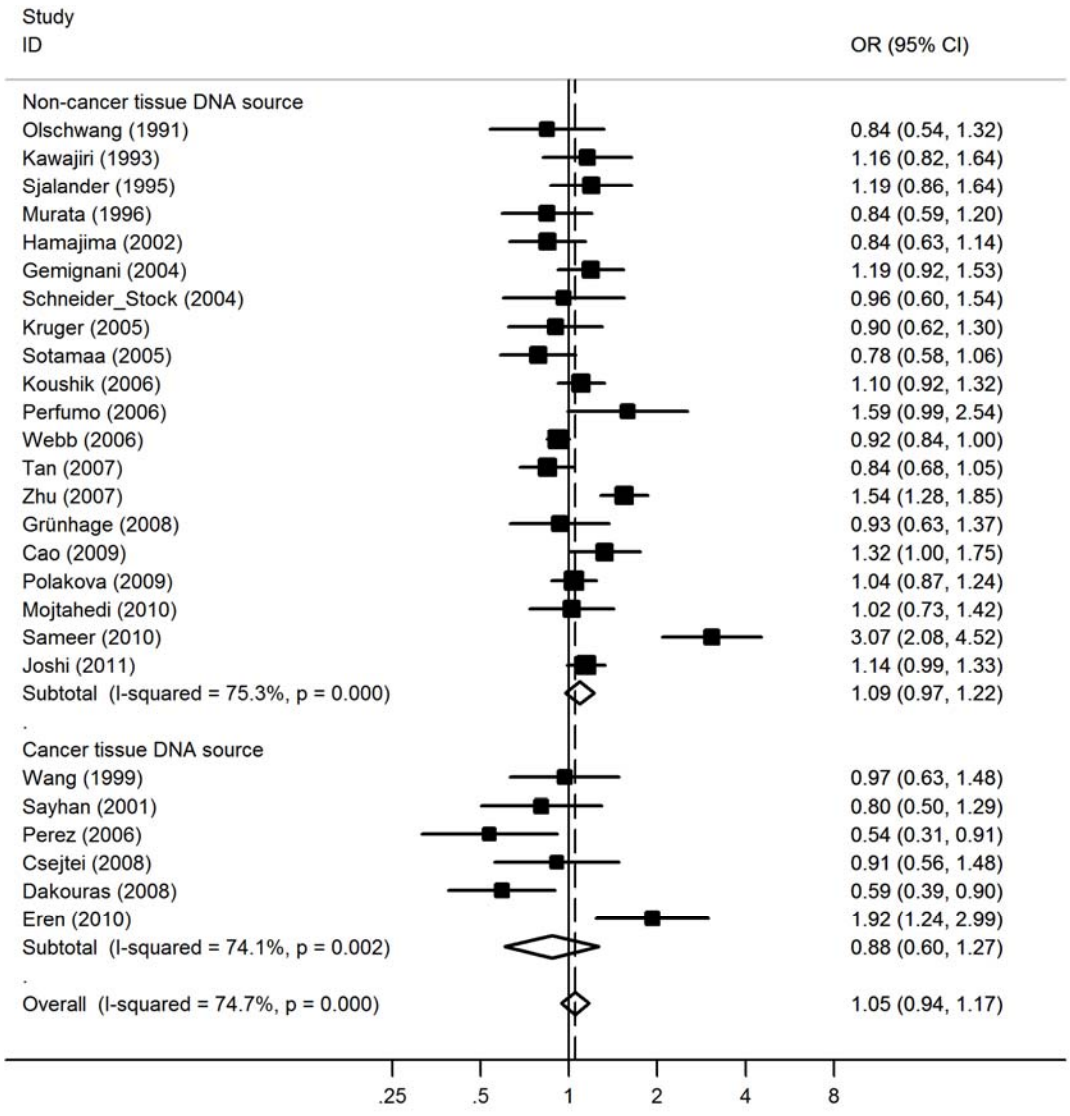
Forest plot for studies in lung cancer. Studies have been stratified by the source of DNA for genotyping cases.



Layout is similar to Supplementary Figure 1.

Appendix Figure 3

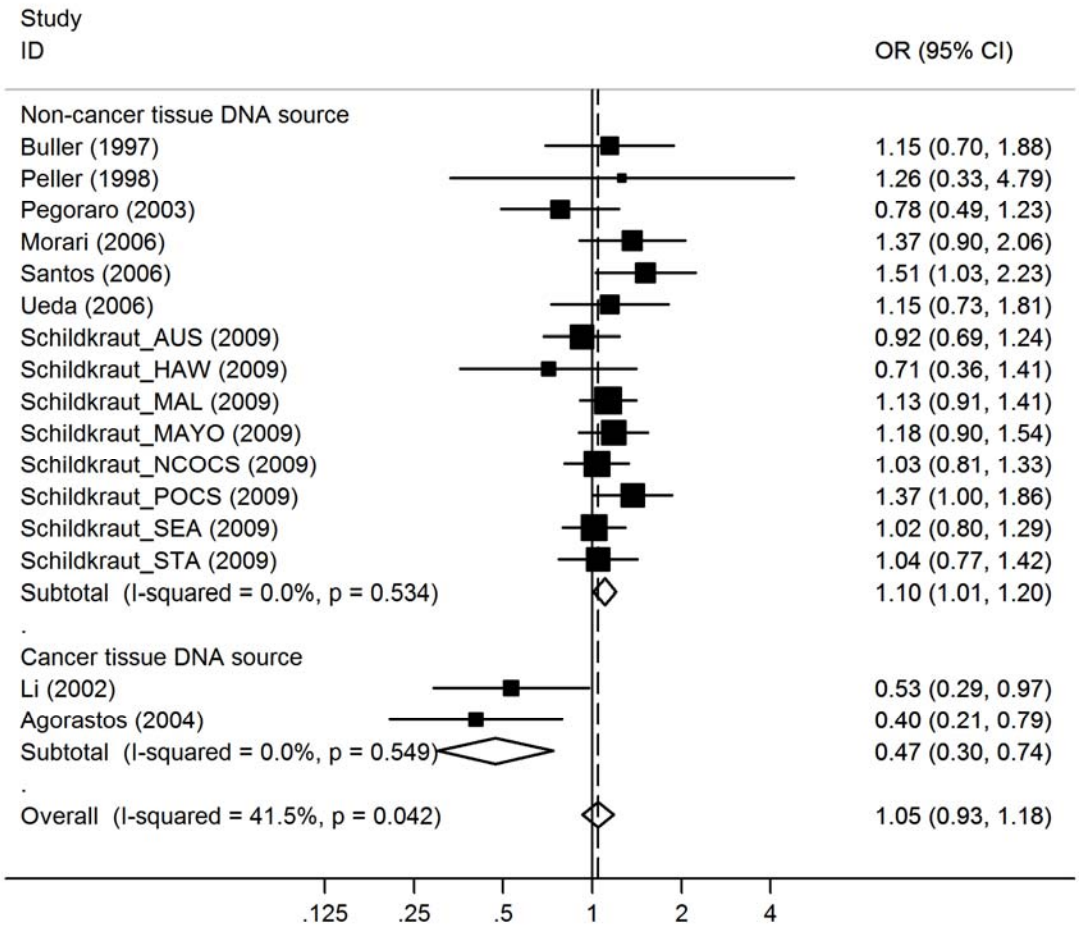
Forest plot for studies in colorectal cancer. Studies have been stratified by the source of DNA for genotyping cases.



Layout is similar to Supplementary Figure 1.

Appendix Figure 4

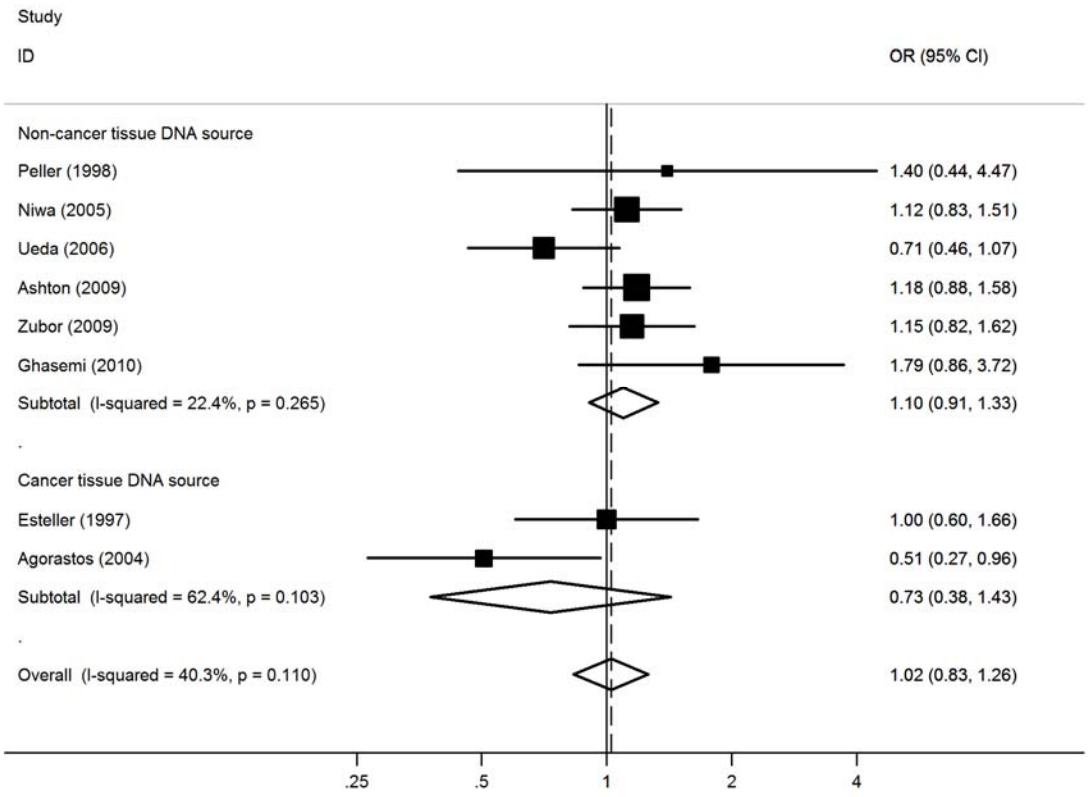
Forest plot for studies in ovarian cancer. Studies have been stratified by the source of DNA for genotyping cases.



Layout is similar to Supplementary Figure 1.

Appendix Figure 5

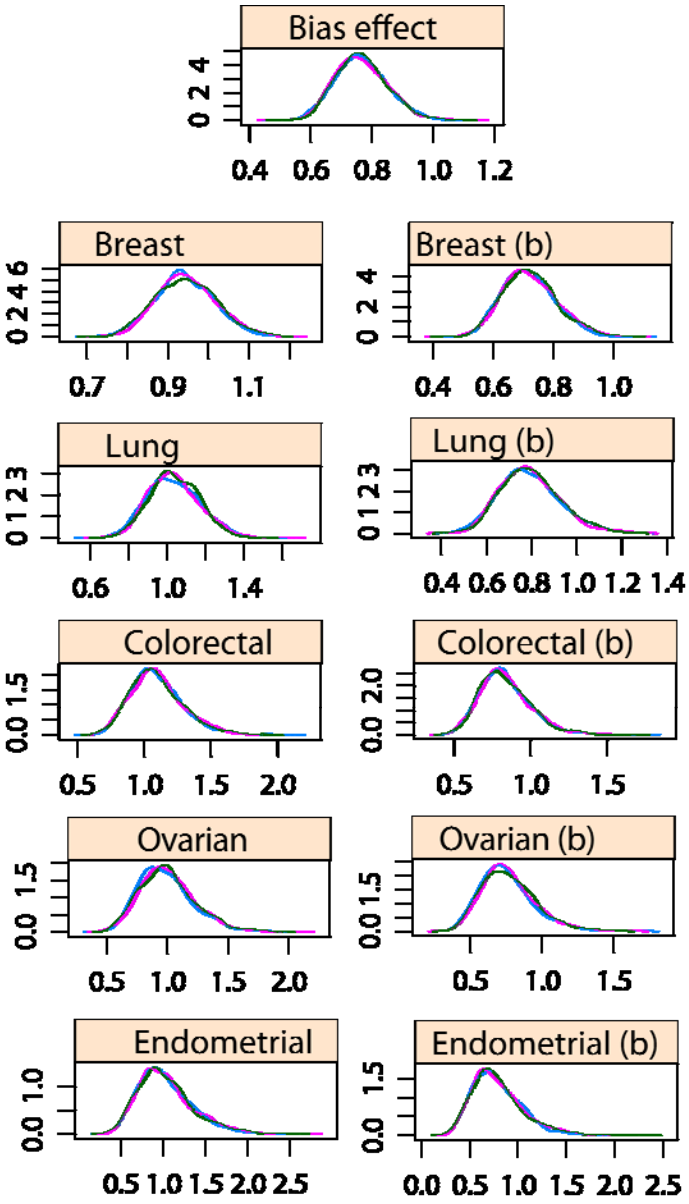
Forest plot for studies in endometrial cancer. Studies have been stratified by the source of DNA for genotyping cases. Layout is similar to Supplementary Figure 1.



Layout is similar to Supplementary Figure 1.

Appendix Figure 6

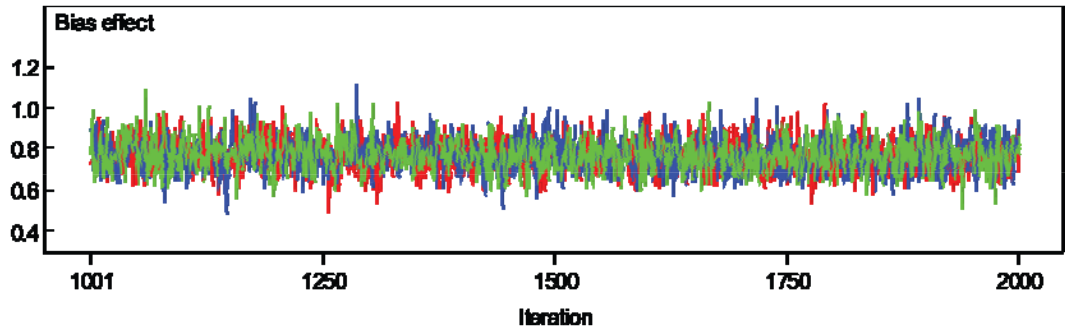
Bayesian posterior distributions for the summary effect sizes from the 2-level Bayesian model using an informative prior.



Posterior probability distributions for selected parameters of interest. The top panel is the posterior distribution of the bias effect. The remaining panels represent posterior distributions of the summary odds ratios for studies in breast, lung, colorectal, ovarian and endometrial cancer using appropriate DNA sources (left) or tumor tissue (indicated with b = bias). Note that distributions for studies using tumor tissue are displaced to the left (compared to the corresponding estimates from studies not using tumor tissue). Three MCMC chains are shown for each parameter (lines in different colors).

Appendix Figure 7

Mixing of chains for the 2-level Bayesian model using an informative prior. Simulation hains for a single parameter (the bias effect are shown).



Trace plot of three chains (different initial values for the Bayesian simulation) for the bias effect across cancers. The lack of a discernible pattern is consistent with successful convergence.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: a cancer journal for clinicians. 2010; **60**(5): 277-300.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA: a cancer journal for clinicians. 2011; **61**(2): 69-90.
3. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000; **343**(2): 78-85.
4. Whibley C, Pharoah PD, Hollstein M. p53 polymorphisms: cancer implications. Nat Rev Cancer. 2009; **9**(2): 95-107.
5. Li FP, Fraumeni JF, Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? Ann Intern Med. 1969; **71**(4): 747-52.
6. Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. Science. 1991; **253**(5015): 49-53.
7. Lane DP. Cancer. p53, guardian of the genome. Nature. 1992; **358**(6381): 15-6.
8. Matlashewski GJ, Tuck S, Pim D, Lamb P, Schneider J, Crawford LV. Primary structure polymorphism at amino acid residue 72 of human p53. Mol Cell Biol. 1987; **7**(2): 961-3.
9. Thomas M, Kalita A, Labrecque S, Pim D, Banks L, Matlashewski G. Two polymorphic variants of wild-type p53 differ biochemically and biologically. Mol Cell Biol. 1999; **19**(2): 1092-100.
10. Marin MC, Jost CA, Brooks LA, Irwin MS, O'Nions J, Tidy JA, et al. A common polymorphism acts as an intragenic modifier of mutant p53 behaviour. Nat Genet. 2000; **25**(1): 47-54.
11. Dumont P, Leu JI, Della Pietra AC, 3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet. 2003; **33**(3): 357-65.
12. Vineis P, Manuguerra M, Kavvoura FK, Guarrera S, Allione A, Rosa F, et al. A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. J Natl Cancer Inst. 2009; **101**(1): 24-36.
13. Storey A, Thomas M, Kalita A, Harwood C, Gardiol D, Mantovani F, et al. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature. 1998; **393**(6682): 229-34.
14. Klug SJ, Ressing M, Koenig J, Abba MC, Agorastos T, Brenna SM, et al. TP53 codon 72 polymorphism and cervical cancer: a pooled analysis of individual data from 49 studies. Lancet Oncol. 2009; **10**(8): 772-84.
15. Tada M, Furuuchi K, Kaneda M, Matsumoto J, Takahashi M, Hirai A, et al. Inactivate the remaining p53 allele or the alternate p73? Preferential selection of the Arg72 polymorphism in cancers with recessive p53 mutants but not transdominant mutants. Carcinogenesis. 2001; **22**(3): 515-7.
16. Schneider-Stock R, Boltze C, Peters B, Szibor R, Landt O, Meyer F, et al. Selective loss of codon 72 proline p53 and frequent mutational inactivation of the retained arginine allele in colorectal cancer. Neoplasia. 2004; **6**(5): 529-35.
17. Brooks LA, Tidy JA, Gusterson B, Hiller L, O'Nions J, Gasco M, et al. Preferential retention of codon 72 arginine p53 in squamous cell carcinomas of the vulva

occurs in cancers positive and negative for human papillomavirus. Cancer Res. 2000; **60**(24): 6875-7.
18. Dahabreh IJ, Linardou H, Bouzika P, Varvarigou V, Murray S. TP53 Arg72Pro polymorphism and colorectal cancer risk: a systematic review and meta-analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010; **19**(7): 1840-7.
19. Schneider-Stock R, Mawrin C, Motsch C, Boltze C, Peters B, Hartig R, et al. Retention of the arginine allele in codon 72 of the p53 gene correlates with poor apoptosis in head and neck cancer. Am J Pathol. 2004; **164**(4): 1233-41.
20. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004; **36**(5): 431-2.
21. Yu W, Yesupriya A, Wulf A, Qu J, Gwinn M, Khoury MJ. An automatic method to generate domain-specific investigator networks using PubMed abstracts. BMC Med Inform Decis Mak. 2007; **7**: 17.
22. Hainaut P, Hernandez T, Robinson A, Rodriguez-Tome P, Flores T, Hollstein M, et al. IARC Database of p53 gene mutations in human tumors and cell lines: updated compilation, revised formats and new visualisation tools. Nucleic Acids Res. 1998; **26**(1): 205-13.
23. Hainaut P, Soussi T, Shomer B, Hollstein M, Greenblatt M, Hovig E, et al. Database of p53 gene somatic mutations in human tumors and cell lines: updated compilation and future prospects. Nucleic Acids Res. 1997; **25**(1): 151-7.
24. Hamroun D, Kato S, Ishioka C, Claustres M, Beroud C, Soussi T. The UMD TP53 database and website: update and revisions. Hum Mutat. 2006; **27**(1): 14-20.
25. Beroud C, Soussi T. The UMD-p53 database: new mutations and analysis tools. Hum Mutat. 2003; **21**(3): 176-81.
26. Zou G, Zhao H. The impacts of errors in individual genotyping and DNA pooling on association studies. Genet Epidemiol. 2004; **26**(1): 1-10.
27. Jawaid A, Sham P. Impact and quantification of the sources of error in DNA pooling designs. Ann Hum Genet. 2009; **73**(1): 118-24.
28. Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. PLoS Med. 2005; **2**(12): e334.
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986; **7**(3): 177-88.
30. Higgins J, Thompson SG, Spiegelhalter DJ. A re evaluation of random effects meta analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2009; **172**(1): 137-59.
31. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. Am J Epidemiol. 1999; **150**(5): 469-75.
32. Cochran W. The Combination of Estimates from Different Experiments. Biometrics. 1954; **10**(1): 101-29.
33. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; **21**(11): 1539-58.
34. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002; **21**(11): 1559-73.

35. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med. 2003; **22**(17): 2693-710.
36. Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata: Stata Corp; 2008.
37. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models: CRC Press; 2004.
38. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random effects meta analysis: A comparative study. Statistics in medicine. 1995; **14**(24): 2685-99.
39. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models: Cambridge University Press Cambridge, UK:; 2007.
40. Abrams K, Myles J, Spiegelhalter D. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley, Chichester, UK; 2004.
41. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical science. 1992; **7**(4): 457-72.
42. Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; **7**(4): 434-55.
43. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. 2nd ed. New York: Chapman and Hall; 2003.
44. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Statistics in medicine. 2006; **25**(20): 3443-57.
45. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997; **315**(7109): 629-34.
46. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ. 2006; **333**(7568): 597-600.
47. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959; **22**(4): 719-48.
48. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990; **1**(1): 43-6.
49. Nelson HH, Wilkojmen M, Marsit CJ, Kelsey KT. TP53 mutation, allelism and survival in non-small cell lung cancer. Carcinogenesis. 2005; **26**(10): 1770.
50. Papadakis E, Soulitzis N, Spandidos D. Association of p53 codon 72 polymorphism with advanced lung cancer: the Arg allele is preferentially retained in tumours arising in Arg/Pro germline heterozygotes. British journal of cancer. 2002; **87**(9): 1013-8.
51. Furihata M, Takeuchi T, Matsumoto M, Kurabayashi A, Ohtsuki Y, Terao N, et al. p53 mutation arising in Arg72 allele in the tumorigenesis and development of carcinoma of the urinary tract. Clinical cancer research. 2002; **8**(5): 1192.
52. Bonafe M, Ceccarelli C, Farabegoli F, Santini D, Taffurelli M, Barbi C, et al. Retention of the p53 codon 72 arginine allele is associated with a reduction of disease-free and overall survival in arginine/proline heterozygous breast cancer patients. Clinical cancer research. 2003; **9**(13): 4860.
53. Kawaguchi H, Ohno S, Araki K, Miyazaki M, Saeki H, Watanabe M, et al. p53 polymorphism in human papillomavirus-associated esophageal cancer. Cancer research. 2000; **60**(11): 2753.

54. Dai S, Mao C, Jiang L, Wang G, Cheng H. P53 polymorphism and lung cancer susceptibility: a pooled analysis of 32 caseñcontrol studies. Human genetics. 2009; **125**(5): 633-8.
55. Matakidou A, Eisen T, Houlston R. TP53 polymorphisms and lung cancer risk: a systematic review and meta analysis. Mutagenesis. 2003; **18**(4): 377.
56. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer. 2005; **5**(2): 142-9.
57. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer. 2004; **4**(4): 309-14.
58. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, et al. Parameters for reliable results in genetic association studies in common disease. Nat Genet. 2002; **30**(2): 149-50.
59. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008; **27**(8): 1133-63.