Review

Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: A meta-analysis

Chikako Kiyohara[1] and Kouichi Yoshimasu[2]

1. Department of Preventive Medicine, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.
2. Department of Hygiene, School of Medicine, Wakayama Medical University, 811-1 Kimiidera, Wakayama 641-8509, Japan

Correspondence to: Chikako Kiyohara, PhD, Department of Preventive Medicine, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. Tel: +81 92 642 6113; Fax: +81 92 642 6115; e-mail: chikako@phealth.med.kyushu-u.ac.jp

Various DNA alterations can be caused by exposure to environmental and endogenous carcinogens. Most of these alterations, if not repaired, can result in genetic instability, mutagenesis and cell death. DNA repair mechanisms are important for maintaining DNA integrity and preventing carcinogenesis. Recent lung cancer studies have focused on identifying the effects of single nucleotide polymorphisms (SNPs) in candidate genes, among which DNA repair genes are increasingly being studied. Genetic variations in DNA repair genes are thought to modulate DNA repair capacity and are suggested to be related to lung cancer risk. We identified a sufficient number of epidemiologic studies on lung cancer to conduct a meta-analysis for genetic polymorphisms in nucleotide excision repair pathway genes, focusing on xeroderma pigmentosum group A (XPA), excision repair cross complementing group 1 (ERCC1), ERCC2/XPD, ERCC4/XPF and ERCC5/XPG. We found an increased risk of lung cancer among subjects carrying the *ERCC2* 751Gln/Gln genotype (odds ratio (OR) = 1.30, 95% confidence interval (CI) = 1.14 - 1.49). We found a protective effect of the *XPA* 23G/G genotype (OR = 0.75, 95% CI = 0.59 - 0.95). Considering the data available, it can be conjectured that if there is any risk association between a single SNP and lung cancer, the risk fluctuation will probably be minimal. Advances in the identification of new polymorphisms and in high-throughput genotyping techniques will facilitate the analysis of multiple genes in multiple DNA repair pathways. Therefore, it is likely that the defining feature of future epidemiologic studies will be the simultaneous analysis of large samples.

Key words: Lung cancer, nucleotide excision repair, meta-analysis, genetic polymorphism

1. Introduction

Sporadic cancer is a multifactorial disease that results from complex interactions between many genetic and environmental factors [1]. This means that there will not be a single gene or single environmental factor that has large effects on cancer susceptibility. Environmental factors (e.g. tobacco smoke, dietary factors, infectious agents and radiation) add to the carcinogenic load to which humans are exposed, but exact numbers for added risk are generally less well established.

Cancer is the result of a series of DNA alterations in a single cell or clone of that cell, which leads to a loss of normal function, aberrant or uncontrolled cell growth and often metastasis. Several of the genes that are frequently lost or mutated have been identified, including genes that function to induce cell proliferation under specific circumstances (e.g. the ras and myc proto-oncogenes) and those which are programmed to halt proliferation in damaged cells (e.g. the *TP53* and *RB1* tumor suppressor genes). Other mutations in genes involved in DNA repair are also necessary. About 150 human DNA repair genes have

been identified to date [2], but the real number is probably higher, since less than 50% of known and putative genes have an identified function. The association between defects in DNA repair and cancer was established by Cleaver in 1968 [3], who showed that xeroderma pigmentosum (XP) is caused by deficient nucleotide excision repair (NER). For more than a quarter of a century after that it was thought that only rare syndromes, such as XP, Cockayne syndrome (CS) and ataxia telangiectasia, were associated with DNA repair defects [4]. Novel, common polymorphisms in DNA repair genes are continuously being identified, and these polymorphisms may play a pivotal role in sporadic carcinogenesis. A growing body of literature, including observations of inter-individual differences in measures of DNA damage, suggests that these polymorphisms may alter the functional properties of DNA repair enzymes.

At least four pathways of DNA repair operate on specific types of damaged DNA. Base excision repair (BER) operates on small lesions, while the NER pathway repairs bulk lesions. Mismatch repair corrects replication errors. Double-strand DNA break repair (DSBR) actually consists of two pathways, homolo-

gous recombination (HR) and non-homologous end-joining (NHEJ). The NHEJ repair pathway involves direct ligation of the two double strand break ends, while HR is a process by which double-strand DNA breaks are repaired through the alignment of homologous sequences of DNA. The following sections review the literature on DNA repair genes in more detail, specifically those involved in the NER pathway.

NER is a versatile DNA repair system that removes a wide range of DNA lesions including UV-induced lesions. There are two subpathways in NER. One is transcription-coupled DNA repair (TCR), which preferentially removes DNA damage that blocks ongoing transcription in the transcribed DNA strand of active genes. The other is global genome repair (GGR), which removes lesions throughout the genome, including those from the nontranscribed strand in the active gene [5]. Three rare, autosomal recessive inherited human disorders are associated with impaired NER activity: XP, CS and trichothiodystrophy (TTD) [6]. XP has been studied most extensively. XP patients develop skin tumors at an extremely high frequency (1000 fold increased incidence as compared to normal individuals) because of their inability to repair UV-induced DNA lesions. These clinical findings are associated with cellular defects, including hypersensitivity to killing and the mutagenic effects of UV and the inability of XP cells to repair UV-induced DNA damage [7]. Approximately 80% of XP patients who have been classified have a defect in the NER pathway. These patients are said to have "classical" XP, in contrast to the remaining 20% of patients who are designated as XP variants (XPV) and most likely have a defect in post-replication repair. In XPV patients, DNA replication stops or is interrupted at sites of UV-damage. Furthermore, *de novo* DNA synthesis opposite cyclobutane pyrimidine dimer lesions is prone to errors, leading to the fixation of multiple DNA mutations and ultimately to cancer. Seven different DNA NER genes, which correct seven distinct genetic XP complementation groups (XPA, XPB (ERCC3), XPC, XPD (ERCC2), XPE, XPF (ERCC4) and XPG (ERCC5, this gene causes CS)) and XPV have been identified [6]. XPA, ERCC3/XPB, ERCC2/XPD, ERCC4/XPF and ERCC5/XPG have a defect in TCR and GGR, while XPC and XPE have a defect in GGR only. ERCC6 and ERCC8 are also known as CS type B (CSB) and CSA, respectively. Approximately 20% of patients have been assigned to the CSA complementation group Essentially CS shows some overlap with certain forms of XP. In contrast to XP and TTD, however, the NER defect in CS is limited to the TCR pathway. As with XP, TTD involves mutations in XP genes, usually *XPD*, which encodes a component of the transcription factor TFIIH [8]. However, it has been suggested that the functions of XPD associated with TTD are distinct from those of XPD associated with XP. Approximately half of the patients with TTD display photosensitivity, correlated with the NER defect.

The aim of this article is to review and evaluate associations between genes in the NER pathway and lung cancer risk, focusing on genes encoding five key enzymes in this pathway: XPA, ERCC1, ERCC2/XPD, ERCC4/XPF and ERCC5/XPG.

2. Materials and methods

2-1. Identification and eligibility of relevant studies

We conducted MEDLINE, Current Contents and Web of Science searches using "XPA", "ERCC1", "ERCC2/XPD", "ERCC4/XPF", "ERCC5/XPG", "lung cancer" and "polymorphism" as keywords to search for papers published (from January 1, 1966 through May 31, 2006). Additional articles were identified through the references cited in the first series of articles selected. Articles included in the meta-analysis were in any language, with human subjects, published in the primary literature and had no obvious overlap of subjects with other studies. We excluded studies with the same data or overlapping data by the same authors. Case-control studies were eligible if they had determined the distribution of the relevant genotypes in lung cancer cases and in concurrent controls using a molecular method for genotyping. Using the MEDLINE database, we identified 5 genetic epidemiological studies [9-13] that provided information on lung cancer occurrence associated with the *XPA* G23A polymorphism (one of the identified 6 candidate studies was excluded due to overlapping data [11]). We identified 5 studies of the *ERCC1* T19007C polymorphism (all of 5 candidate studies were independent [13-17]). We gathered 18 articles on the *ERCC2* 312/751 polymorphisms found through literature searches and checked their references for additional relevant studies. Of the relevant 18 studies, 2 studies appeared to be on populations already reported [14, 18, 19], leaving 15 independent studies (11 studies for the Asp312Asn polymorphism [11, 13, 14, 17-24] and 14 studies for the Lys751Gln polymorphism [11, 13, 14, 17-19, 21-28]. Less than 5 studies each have been reported on the *ERCC1* C8092A, *ERCC4/XPF* Arg415Gln, *ERCC4/XPF* Ser835Ser, *ERCC5/XPG* His46His, *ERCC5/XPG* Asp1104His SNPs.

2-2. Data extraction and assessment of study quality

For each study, characteristics such as authors, year of publication, ethnic group of the study population, source of control population, number of genotyped cases and controls, crude odds ratio (OR) and the method for quality control of genotyping were noted. For studies including subjects of different ethnic groups, data were extracted separately for each ethnic group whenever possible.

Methods for defining study quality in genetic studies are more clearly delineated than those for observational studies. We assessed the homogeneity of the study population (Caucasian or Asian).

2-3. Meta-analysis

Data were combined using both a fixed effects (the inverse variance-weighted method) and a random effects (DerSimonian and Laird method) models [29].

The Cochrane Q statistics test is used for the assessment of heterogeneity. The fixed effects model is used when the effects are assumed to be homogenous, while the random effects model is used when they are heterogenous. In the absence of between-study heterogeneity, the two methods provide identical results. The presence of heterogeneity can result from differences in the selection of controls, age distribution, prevalence of lifestyle factors, histologic type of lung cancer, stage of lung cancer and so on. The random effects model incorporates an estimate of the between-study variance and tends to provide wider CIs when the results of the constituent studies differ among themselves. As the random effects model is more appropriate when heterogeneity is present [29], the summary OR and prevalence were essentially based on the random effects model. The meta-analyses were performed on crude ORs, since the adjusted ORs were not comparable because of the inclusion of different covariates in the multivariate regression models. Using individuals with the homozygous common genotype as the reference group, we calculated ORs for individuals with the heterozygous genotype and homozygous rare genotype separately whenever possible (information available in at least two studies). In some cases, we combined the heterozygous genotype with the homozygous rare genotype due to a low prevalence of the rare allele in several polymorphisms. The Q statistic was considered significant for P<0.10 [30, 31]. Publication bias is always a concern in meta-analysis. The presence of publication bias indicates that nonsignificant or negative findings remain unpublished. To test for publication bias, both Begg's [32] and Egger's [33] tests are commonly used to assess whether smaller studies reported greater associations than larger studies. Publication bias is considered significant for P<0.10. Publication bias may be always a possible limitation of combining data from various sources as in a meta-analysis. The idea of adjusting the results of meta-analyses for publication bias and imputing "fictional" studies into a meta-analysis is controversial at the moment [34]. Sutton et al. concluded that publication or related biases did not affect the conclusions in most meta-analyses because missing studies changed the conclusions in less than 10% of meta-analyses [34]. All of the calculations were performed using STATA Version 8.2 (Stata Corporation, College Station, TX) software.

3. Results

3-1. DNA repair capacity and lung cancer risk

Cigarette smoke contains several thousand chemicals that are known to chemically modify DNA [35] and lead to the formation of mutations [36]. Most of these compounds are procarcinogens that must be activated by Phase I enzymes, such as cytochrome P450s. All activated carcinogens can bind to DNA and form DNA adducts that are capable of inducing mutations and initiating carcinogenesis. The capacity to repair DNA damage induced by activated carcinogens appears to be one of the host factors that may influence lung cancer risk. A critical cellular response that counteracts the carcinogenic effects of DNA damage is DNA repair. As stated earlier, there are several known pathways of DNA repair, all of which act to remove DNA lesions and prevent mutations, thereby restoring genetic integrity.

Several studies have investigated whether reduced DNA repair capacity (DRC) is associated with an increased risk of cancer [37]. The reduced DRC of benzo(a)pyrene-7,8-diol-9,10-epoxide (an active form of benzo(a)pyrene)-DNA adducts is associated with an increased risk of lung cancer (2.1-fold, 95% confidence interval (CI) = 1.5 - 3.0) [38]. The reduced DRC has been shown to be associated with a 5.7-fold (95% CI = 2.1 - 15.7) increased risk of developing lung cancer [39]. Likewise, the reduced DRC of bleomycin-induced damage was found to be associated with an increased risk of lung cancer [40]. These studies suggested that a low DRC of various DNA repair mechanisms predisposes individuals to lung cancer, and this realization prompted us to search for defined DNA repair activities that may be risk factors for lung cancer. Polymorphisms in DNA repair genes may be associated with differences in the DRC of DNA damage and may influence an individual's risk of lung cancer, because the variant genotype in those polymorphisms might destroy or alter repair function.

3-2. *XPA* G23A polymorphism and lung cancer risk

The heterotrimeric replication protein A (RPA) is required for NER and may play an important role in the damage recognition process. The XPA protein is required for NER and is involved in the DNA damage recognition process. Both RPA and XPA preferentially bind damaged DNA, and because RPA and XPA directly interact in the absence of DNA, the RPA-XPA complex has been implicated as a key component in the earliest stage of damage recognition [41]. There is also evidence that the XPC-hHR23B protein complex may initiate recognition of DNA damage for the global genomic repair pathway of NER [42]. Recent evidence also implicates the damaged DNA binding protein heterodimer in damage recognition, because the complex binds damaged DNA with high affinity [43] and can dramatically increase the repair rate of certain DNA adducts, including cyclobutane pyrimidine dimers, in conjunction with XPA and RPA [44].

The *XPA* maps on chromosome 9, at 9q22.3. In the *XPA* gene, a polymorphic site was identified that was in the 5' untranslated region (UTR) of the gene and which consisted of a G-to-A (or A-to G) substitution in the fourth nucleotide before the ATG start codon (dbSNP rs 1800975) [45]. SNP alleles with higher frequencies are more likely to be ancestral than less frequently occurring alleles although there may be some exceptions. As the 23G allele was more prevalent than the 23A allele (Table 1), we regarded the 23G allele as ancestral (wild-type or major) allele for descriptive purposes (the *XPA* 23 polymorphism caused by the G-to-A substitution is the *XPA* G23A poly-

morphism). The polymorphism, termed the *XPA* G23A polymorphism (at position 23 in the transcript, four nucleotides upstream of the start codon), is in the Kozak sequence near the start codon and thus may affect the XPA protein levels in cells [46]. A functional association between the *XPA* G23A polymorphism and DRC has been reported [10]. It has been shown that healthy subjects with at least one 23G allele have significantly higher DRC. When the combined A/A and A/G genotype was used as the reference, the G/G genotype was associated with a significantly decreased risk of lung cancer (adjusted OR = 0.56, 95% CI = 0.35 - 0.90) in Koreans [9]. A significant protective effect of the combined G/A and G/G genotypes on lung cancer risk was reported in Americans (adjusted OR = 0.69, 95% CI = 0.53 - 0.90) and Mexican-Americans (adjusted OR = 0.32, 95% CI = 0.12 - 0.83) [10]. Likewise, a protective and nonsignificant effect was seen among Germans [11] and Danes [12]. As compared with the combined G/A and A/A genotypes, the G/G genotype was, however, associated with a significantly increased risk of lung cancer (adjusted OR = 1.59, 95% CI =1.12 - 2.27) in a Norwegian population [13]. Summary frequencies of the 23A allele among all and Caucasian populations, based on the random effects model, were 0.368 (95% CI = 0.308 - 0.429) and 0.352 (95% CI = 0.277 - 0.428), respectively (Table 1). Summary ORs for the G/A genotype and G/G genotype among 5 studies in 7 populations were 0.73 (95% CI = 0.61 - 0.89) and 0.75 (95% CI = 0.59 - 0.95), respectively (Table 1). Evidence for heterogeneity was absent in both analyses. Among Caucasian studies, the summary ORs for the G/A genotype and the A/A genotype were 0.72 (95% CI = 0.58 - 0.89) and 0.82 (95% CI = 0.61 - 1.11), respectively. The Cochrane Q test for heterogeneity did not show a statistical significance. The Egger's test was statistically significant for publication bias in a subgroup analysis of Caucasians (P = 0.073, G/A genotype vs. G/G genotype).

Two studies investigated associations between cigarette smoking and the G23A polymorphism in relation to lung cancer. When stratifying by smoking status, there was a significant protective effect for current smokers who possessed the G/G genotype (adjusted OR = 0.23, 95% CI = 0.07- 0.71) but not for former or never smokers [9]. Ever smokers (current and former) with at least one copy of the 23G allele showed a significantly reduced risk of lung cancer (adjusted OR = 0.68, 95% CI = 0.51 - 0.91) among Caucasians [10]. The presence of the 23A polymorphism, however, was associated with a statistically significant reduced risk in subjects who smoked >29 pack-years (OR = 0.53, 95% CI = 0.17 - 0.97) [13]. Interactions between cigarette smoking and the polymorphism were not determined in the studies [9, 10, 13]. No associations were seen between the G23A polymorphism and any histologic types of lung cancer [11], while the G/G genotype was associated with a significantly decreased risk for small cell lung cancer (OR = 0.23, 95% CI = 0.07 - 0.71) [9].

The *XPA* G23A polymorphism may, thus, be a promising SNP for lung cancer. It is thought that cigarette smoking modifies the association between DNA repair polymorphisms, as well as metabolic polymorphisms, and lung cancer risk. Since interactions between the G23A polymorphism and smoking have not been fully elucidated, further studies are needed to better understand the associations between the *XPA* G23A polymorphism and lung cancer risk.

3-3. *ERCC1* polymorphisms and lung cancer risk

The *ERCC1* coding region is 1.1 kb long and comprises 10 exons. This gene is located on 19q13.2 - q13.3. Shen et al. [47] have identified polymorphisms of three of the exons of the *ERCC1* gene, all of which resulted in silent mutations. No amino acid substitutions were observed among the *ERCC1* polymorphisms [48]. The functional effects of the silent polymorphisms in *ERCC1* have not been fully elucidated; however, some of the variant alleles of the polymorphisms in DNA repair genes may be associated with the reduced DRC. The studies have focused on polymorphisms of the 3′ UTR (C8092A, dbSNP no. rs3212986) and codon 118 (Asn118Asn, T19007C, dbSNP no. rs11615) in *ERCC1*.

For the T19007 C (Asn118Asn) polymorphism, although the T/T genotype generates the less commonly associated triplet codon sequence encoding the amino acid and has been termed the "variant" by convention, the T/T genotype indeed has been reported to occur at higher frequencies. Hence, the T/T genotype is used as reference in this paper. The C/C genotype of the C8092A polymorphism is used as reference on the same score.

The C/C genotype of the T19007C polymorphism was associated with a significantly decreased risk of lung cancer (adjusted OR = 0.32, 95% CI = 0.19 - 0.55) in a Norwegian population [13]. A lack of association between the T19007C polymorphism and lung cancer risk was observed in a Danish population [14], a large American population [15], a Chinese population [16] and a nonsmoking European population [17]. As shown in Table 2, summary frequencies of the 19007T allele among all and Caucasian populations, based on the random effects model, were 0.499 (95% CI = 0.387 - 0.611) and 0.575 (95% CI = 0.529 - 0.622), respectively. The summary ORs for the T/C genotype and the C/C genotype were 0.82 (95% CI = 0.62 - 1.08) and 0.72 (95% CI = 0.46 - 1.11), respectively. Even if the analysis was restricted to Caucasian studies, the ORs did not materially change. The Cochrane Q test for heterogeneity showed a statistical significance in any analysis. In comparison of the T/C genotype with the T/T genotype, the Begg's test was statistically significant in an overall analysis (P = 0.086) and a subgroup analysis of Caucasians (P = 0.089).

Two studies examined an interaction between the T19007C polymorphism and cigarette smoking. When stratified by smoking status, the interaction between smoking and the polymorphism was not statistically significant [15, 16]. Only one study provides information on the T19007C polymorphism and lung cancer risk in histologic types. There was no difference

in risk estimates according to the histological type of lung cancer [16].

As for the C8092A polymorphism, no association was found between the polymorphism and lung cancer risk in Norwegians [13] and Americans [15]. The C8092A and T19007C polymorphisms have been reported to be in linkage disequilibrium [15].

Although harboring at least one 19007C allele may be associated with a deceased risk of lung cancer, the protective effect of the 19007C allele needs to be confirmed in other independent studies. Furthermore, additional studies are needed to detect the function of the *ERCC1* polymorphisms.

3-4. *ERCC2/XPD* polymorphisms and lung cancer risk

The ERCC2/XPD protein plays a role in the NER pathway, which recognizes and repairs a wide range of structurally unrelated lesions such as bulky adducts and thymidine dimers. ERCC2/XPD works as an ATP-dependent $(5'\rightarrow3')$ helicase joined to the basal TFIIH complex used to separate the double helix. The ERCC2/XPD protein is necessary for normal transcription initiation and NER. *ERCC2/XPD* maps on chromosome 19, at 19q13.3 and covers 21.14 kb. Mutations in the *ERCC2* gene can diminish the activity of TFIIH complexes, giving rise to repair defects, transcription defects and abnormal responses to apoptosis [49].

A number of polymorphisms in the *ERCC2/XPD* gene have been reported. Whereas polymorphisms in the codons 199, 201 and 575 are rare, those in codons 156, 312, 711 and 751 are common. Two *ERCC2/XPD* polymorphisms, Asp312Asn (db SNP no. rs1799793) and Lys751Gln (db SNP no. rs13181), have mainly been investigated in relation to phenotypic endpoints relevant to lung carcinogenesis. With regard to the Asp312Asn polymorphism, most of the reported data indicate a higher level of DNA adducts in subjects with the Asn allele. The interpretation of this finding is a lower DRC for the Asn allele than the Asp allele. This is also true for the *ERCC2/XPD* Lys751Gln polymorphism. The Gln allele is associated with a higher DNA adduct level or lower DRC.

The Asp/Asp genotype of the *ERCC2/XPD* Asp312Asn polymorphism was found to have an increased risk of lung cancer when the combined Asp/Asn and Asn/Asn genotypes served as reference (OR = 1.86, 95% CI =1.02 - 3.40) in Polish men [20]. A large American lung-cancer study also reported an elevated risk (adjusted OR = 1.5, 95% CI = 1.1 - 2.0; Asn/Asn genotype vs. Asp/Asp genotype) [18]. Likewise, Chinese subjects homozygous for the Asn/Asn genotype had an increased risk of lung cancer (adjusted OR = 10.33, 95% CI = 1.29 - 82.50) compared with subjects homozygous for the Asp/Asp genotype [19]. No association with this polymorphism was seen in an admixed population [21], a small Swedish population [22] and among Finnish smoking men [23]. Two meta-analyses have been published in 2004 [50] and 2005 [51], respectively. Both of them are based on the same published data from 6 individual case-control studies [18-23]. The first meta-analysis showed that individuals with the Asn/Asn genotype had a 27% (95% CI = 1.04 - 1.56) increased risk of lung cancer compared with individuals with the Asp/Asp genotype. The results supported the hypothesis that individuals with the Asn/Asn genotype are at higher risk of developing lung cancer [50]. The second meta-analysis was somewhat different from the first one, because unadjusted ORs were summarized in the first one. The summary OR associated with the Asn/Asn genotype was 1.18 (95% CI = 0.84 - 1.67). No significant association between the *ERCC2/XPD* Asp312Asn polymorphism and lung cancer was found in the second meta-analysis [51]. Regardless, these meta-analyses indicate that the excess lung cancer risk from the Asn/Asn genotype may be less than 30%.

Five studies have been reported since the publication of these two meta-analyses. They revealed that the Asp312Asn polymorphism was not associated with lung cancer risk in Germans [11], Norwegians [13], Danes [14], Europeans [17] and Chinese [24].

As shown in Table 3, the summary frequency of the 312Asp allele among Caucasians (0.645, 95% CI = 0.572 - 0.719) was significantly lower than that among Asians (0.936, 95% CI = 0.925 - 0.946). Summary ORs associated with the *ERCC2/XPD* Asp312Asn polymorphism are also shown in Table 3. No significant association between lung cancer and the heterozygous Asp/Asn genotype was found for all of the studies combined or by ethnicity. The Cochrane Q test for heterogeneity did not show a statistical significance in all analyses. Although no evidence of publication bias was found in overall analyses, both Begg's (P= 0.035) and Egger's (P = 0.003) tests showed a statistical significance in a subgroup analysis of Caucasians (Asn/Asn genotype vs. Asp/Asp genotype).

When stratifing by smoking dose, the risk of lung cancer was significantly higher in light-smokers with the Asp/Asp genotype than in those with the Asn/Asn genotype [20]. Similar findings were not seen for never-smoker or heavy-smokers [20]. A significant interaction between smoking (smoking status, pack-years and duration) and the polymorphism was observed in one study [18] but not in two other studies [16,19]. Stratification analysis revealed that the increased risk was mainly confined to squamous cell carcinoma of the lung, with the ORs being 20.50 (95% CI = 2.25 - 179.05) for the 312Asn/Asn genotype [19].

Table 4 shows the association between the *ERCC2* Lys751Gln polymorphism and lung cancer risk. The Gln/Gln genotype was associated with an increased risk for lung cancer compared with the 751Lys/Lys genotype (adjusted OR = 2.71, 95% CI = 1.01 - 7.24) in Chinese [19]. Stratification analysis revealed that the increased risk was mainly confined to lung squamous cell carcinoma, with the OR being 4.24 (95% CI = 1.34 - 13.38) for the Gln/Gln genotype [19], however. Although David-Beades et al. reported that the Gln/Gln genotype was associated with a significantly increased risk of lung cancer in Caucasians

(USA), a multivariate-adjusted OR was no longer significant [25]. No association with the Lys751Gln polymorphism was seen in two Caucasian populations [18, 22], an admixed population [21], a Finnish population [23], African-Americans [25], a Chinese population [26] and a Korean population [27]. The meta-analysis by Hu et al. (2004) showed that the Gln/Gln genotype had a 21% (95% CI = 1.02 - 1.43) increased risk of lung cancer compared with individuals with the Lys/Lys genotype [51]. The meta-analysis by Benhamou and Sarasin (2005) reported that the summary OR for the Gln/Gln genotype was 1.18 (95% CI = 0.95 - 1.47) [51]. Both of the meta-analyses were based on the same published data from 8 individual case-control studies [18, 19, 21-23, 25-27]. No significant association between the Lys751Gln polymorphism and lung cancer was found in the two meta-analyses [51]. These meta-analyses indicate that the excess lung cancer risk from the Gln/Gln genotype may be about 20%. Six studies [11, 13, 14, 17, 24, 28] have been reported after the two meta-analysis. Danish subjects with the Gln/Gln genotype were at a 2.01-fold (95% CI = 1.20 - 3.35) higher risk of lung cancer risk than those with the Lys/Lys genotype [14]. Similarly, the Gln/Gln genotype was associated with significantly increased risk of lung cancer (adjusted OR = 1.60, 95% CI = 1.10 - 2.30) in Norwegians [13]. German individuals with the Gln/Gln genotype were at a borderline increased risk (adjusted OR = 1.59, 95% CI = 0.95 - 2.67) [11]. However, individuals with the Gln allele had a 61% (95% CI = 14 - 83) reduction of lung cancer risk in a Chinese population [24]. No association with the Lys751Gln polymorphism was seen in a European cohort [17] and in non-Hispanic Caucasians (USA) [28].

The summary frequency of the 751Lys allele among Caucasians (0.634, 95% CI = 0.614 - 0.655) was significantly lower than that among Asians (0.843, 95% CI = 0.763 - 0.924). A statistically significant ethnic difference was observed between Caucasians and Asians. Summary ORs for the Gln/Gln genotype and Lys/Gln genotype were 1.06 (95% CI = 0.97 - 1.16) and 1.30 (95% CI = 1.14 - 1.49), respectively. Evidence of publication bias was absent in all of the analyses. The effect of the Gln/Gln genotype on lung cancer risk was stronger in Caucasians (OR = 2.25, 95% CI = 0.97 - 5.23) than in Asians (OR = 1.02, 95% CI = 0.20 - 5.27). This may only be due to a difference in sample sizes. Reasons for this difference in risk among different ethnic populations are as yet unknown but, if real, may be related to other genetic or environmental factors. The Cochrane Q test for heterogeneity showed a statistical significance among Asian studies (P = 0.040, Gln/Gln genotype vs. Lys/Lys genotype).

There was no interaction between smoking (smoking status, pack-years and duration) and the polymorphism [14, 19, 26, 27]. Although the Lys/Lys genotype was associated with a statistically significant increased risk (OR = 2.0, 95% CI = 1.15 - 3.41) among subjects who smoked>29 pack-years, an interaction between cigarette smoking and the polymorphism

was not determined [13]. When stratified by histological type, no statistically significant association between the polymorphism and lung cancer risk was found [26, 27].

Several studies have investigated the possible association of *ERCC2/XPD* Asp312Asn and Lys751Gln polymorphisms with lung cancer with inconsistent results. The Lys751Gln polymorphism has been more studied than the Asp312Asn polymorphism, because the frequency of the 751Gln allele is more prevalent than the 312Asn allele. The Asp312Asn polymorphism is in linkage disequilibrium with the Lys751Gln polymorphism [19, 20, 21], however. The inconsistent associations in previous studies of the *ERCC2/XPD* polymorphisms could be due to differences in study populations, the small sample sizes of earlier studies and possible environmental interactions.

3-5. *ERCC4/XPF* polymorphisms and lung cancer risk

ERCC4/XPF is an essential protein in the NER pathway, which is responsible for removing UV-C photoproducts and bulky adducts from DNA. Among the NER enzymes, ERCC4/XPF and ERCC1 are also uniquely involved in removing DNA interstrand cross-linking damage. The ERCC4/XPF-ERCC1 complex, which makes incisions at the 5′ end of DNA loops, may contribute to the repair of large trinucleotide repeat containing loops that are generated due to replication slippage and that are too long to be repaired by the postreplicative DNA mismatch repair system [52]. Polymorphisms in enzymes involved in large loop repair could be responsible for the observed variation in the stability of similar-sized trinucleotide repeat disease alleles among different individuals. The *ERCC4/XPF* gene is evolutionarily conserved. Extensive homology exists between human ERCC4/XPF, *Drosophila* Mei-9, *Saccharomyces cerevisiae* RAD1, and *S. pombe* Rad16 [53], all of which have similar functions in NER.

The *ERCC4/XPF* gene contains 11 exons, spans 28.2 kb and is located on chromosome 16p13.2 - p13.13. Several polymorphisms exist in the coding region of *ERCC4/XPF*, a few of which have been associated with cancer risks. Genetic instability of simple repeated sequences might also be influenced by the *ERCC4/XPF* polymorphisms. The *ERCC4/XPF* G1244A polymorphism is a G-to-A change in exon 8 (Arg415Gln, dbSNP no. rs1800067) that results in a change from arginine to glutamine. The *ERCC4/XPF* polymorphism in exon 8 has been reported to be associated with an increased risk for developing breast cancer [54]. The T2505C polymorphism is a T-to-C change in exon 11 (Ser835Ser, dbSNP no. rs1799801) that results in no amino acid change (serine is conserved) [55]. Functionally significant SNPs in the *ERCC4/XPF* gene may also contribute to individual differences in the fine details of DNA repair. A lack of association was found between the G1244A (Arg415Gln) polymorphism and lung cancer risk (adjusted OR = 1.11, 95% CI = 0.59 - 2.07; Arg/Gln genotype vs. Arg/Arg genotype) in

Koreans [9]. The C/C genotype of the T2505C polymorphism was nonsignificantly associated with an increased risk of lung cancer (adjusted OR = 1.71, 95% CI = 0.52 - 5.58) in Chinese [24].

3-6. *ERCC5/XPG* polymorphisms and lung cancer risk

ERCC5/XPG is responsible for a 1186 amino acid structure-specific endonuclease activity that is essential for the two incision steps in NER. The ERCC5/XPG nuclease has been suggested to act on the single-stranded region created as a result of the combined action of the XPB helicase and the ERCC2/XPD helicase at the DNA damage site. In human cells, ERCC5/XPG catalyses an incision approximately 5 nucleotides 3' to the site of damage but is also involved non-enzymatically in the subsequent 5' incision. It is further involved in the stabilization of a pre-incision complex on the damaged DNA.

The *ERCC5/XPG* gene contains 17 exons, spans 32 kb and is located on chromosome 13q32.3 -q33.1. Several polymorphisms in the coding sequence of the *EECC5/XPG* gene have been identified. The association between lung cancer and two common polymorphisms, T335C (His46His, dbSNP no. rs1047768) and G3507C (Asp1104His, dbSNP no. rs17655), have been investigated. The functional effects of these two SNPs are still unknown. However, it is likely that the SNPs in the coding DNA sequences may result in a subtle structural alteration of the ERCC5/XPG activity and modulation of lung cancer susceptibility.

The Asp/Asp genotype of the Asp1104His polymorphism was associated with a significantly decreased risk of lung cancer (adjusted OR = 0.60, 95% CI = 0.38 - 0.95) in a Korean population [56]. Similarly, the Asp/Asp genotype was inversely associated with lung cancer (adjusted OR = 0.65, 95% CI = 0.39 - 1.1) in an admixed population (composed mostly composed of whites) [57]. However, the Asp/Asp genotype was not associated with lung cancer risk in a Chinese population [24]. As for T335C polymorphism, the C/C genotype was associated with a significantly increased risk of lung cancer (adjusted OR = 1.79, 95% CI = 1.19 - 2.63) in Norwegians [13] but not in Chinese [24].

4. Discussion

Epidemiological studies of common polymorphisms in DNA repair genes, if large and unbiased, can provide insight into the *in vivo* relationships between DNA repair genes and lung cancer risk. Such studies may identify empirical associations which indicate that a polymorphism in a gene of interest has an impact on lung cancer, independent of metabolic regulatory mechanisms and other genetic and environmental variability. Findings from epidemiological studies can complement *in vitro* analyses of the various polymorphisms, genes, and pathways. In addition, epidemiological studies of common polymorphisms can lead to an increased understanding of the public health dimension of DNA-repair variation.

We conducted a systematic literature review to evaluate the associations between sequence variants in DNA repair genes and lung cancer risk. We found an increased risk of lung cancer among subjects carrying the *ERCC2/XPD* 751Gln/Gln genotype (OR = 1.30, 95% CI = 1.14 - 1.49). The Gln allele of the *ERCC2/XPD* Lys751Gln polymorphism is associated with a higher DNA adduct level or lower DNA repair efficiency, except in research published by Duell et al. (2000) who found no correlation between the *ERCC2/XPD* Lys751Gln polymorphism and the level of polyphenol-DNA adducts in human blood samples [58]. Matullo et al. (2003) demonstrated a higher level of DNA adducts, measured by ^{32}P-postlabeling, in lymphocytes from nonsmokers with the *ERCC2/XPD* 751Gln/Gln genotype [59]. Similarly, Palli et al. (2001) reported a higher level of DNA adducts in workers with at least one Gln allele who were exposed to traffic pollution in comparison with workers with the two common alleles [60]. An increased number of aromatic DNA adducts was found by Hou et al. (2002) in peripheral blood lymphocytes from subjects with the *ERCC2/XPD* 312Asn and *ERCC2/XPD* 751Gln alleles [22]. The combined Asn/Asn and Gln/Gln genotypes showed a higher level of DNA lesions than did other genotypes.

In contrast, we found a protective effect of the *XPG* G23A G/G genotype (OR = 0.75, 95% CI = 0.59 - 0.95) on lung cancer risk. The G23A polymorphism itself may alter the transcription and/or translation of the gene. Because this polymorphism is located in the vicinity of the translation initiation codon, it may alter translation efficiency. The nearby proximal nucleotides to the AUG initiation codon are important for the initiation of translation because the 40S ribosomal subunit binds initially at the 5'-end of the mRNA [61]. The consensus sequence around the start codon is GCCRCCAUGG, which is known as the Kozak consensus sequence [62]. The R at position -3 and the G just downstream of the start codon are especially important, and the lack of these bases leads to read-through of the start codon [63]. However, there has been no precise explanation of the mechanism by which the recognition of the start codon is aided by a purine at position -3 [62], which is the core nucleotide of the Kozak consensus. The polymorphism *XPA* G23A is a G/A transversion occurring 4 nucleotides upstream of the start codon of *XPA* and possibly improving the Kozak sequence [9]. The sequences (CCAGAGAUGG) around the predicted initiator methionine codon of the *XPA* gene agree with the Kozak's consensus sequence at positions -3 and +4 [64]. Although both the A and polymorphic variant G nucleotides at the -4 position of the *XPA* gene do not correspond to the original consensus Kozak sequence containing the nucleotide C at position -4, it is possible that a nucleotide substitution of A to G at position -4 preceding the AUG codon may affect ribosomal binding and thus alter the efficiency of XPA protein synthesis. To investigate whether the transition from G to A changes the translation efficiency, an *in vitro* transcription/translation analysis and a primer extension assay of the initiation complex will be necessary in the

future. Furthermore, a functional association between the G23A polymorphism and DRC was reported [10], which showed significantly higher repair efficiency in healthy subjects with at least one G allele. An alternative explanation could be that the protective *XPA* 23G allele is in linkage disequilibrium with an allele from an adjacent gene which is the true susceptibility gene.

Several DNA repair pathways are involved in the maintenance of genetic stability. The most versatile and important one is the NER pathway, which detects and removes bulky DNA adducts, including those induced by cigarette smoking [65]. However, there are several conflicting reports on the association between this polymorphism and lung cancer risk among various populations. Although the reasons for the inconsistencies in the studies are not clear, possible explanations are: 1) low frequency of the "at-risk" genotype, which would reduce the statistical power of the studies and 2) small size of the studies. Ethnic differences in the roles of the polymorphism may be caused by gene-gene interactions, different linkages to the polymorphisms determining lung cancer risk and different lifestyles.

The most important problems facing lung cancer research are identifying "at-risk" individuals and implementing clinical surveillance, prevention practices, and follow-up care. Repair pathways play an important role in lung cancer risk, and genetic variations may contribute to decreased DRC and lung cancer susceptibility. Although the increased/decreased risk associated with individual DNA repair SNPs may be small compared to that conferred by high-penetrance cancer genes, their public health implication may be large because of their high frequency in the general population. It is thus essential that epidemiological investigations of DNA repair polymorphisms are adequately designed. Unfortunately a fairly good number of studies are limited by their sample size and subsequently suffer from too low power to detect effects that may truly exist. Also, given the borderline significance of some associations and multiple comparisons that have been carried out, there is a possibility that one or more findings are false-positives [66]. Large and combined analyses may be preferred to minimize the likelihood of both false-positive and false-negative results. In addition, controls should be chosen in such a way that, if they were cases, they would be included in the case group; when controls are matched to cases, it is essential to account for matching in the analysis. When appropriate, confounding factors should be controlled for, with particular consideration of race and ethnicity. An additional major concern is the grouping of genotypes for calculation of ORs. Without functional data to dictate genotype groupings, it seems prudent to present two ORs per polymorphism (one for heterozygotes vs. common-allele homozygotes and one for rare-allele homozygotes vs. common-allele homozygotes) so that dominant, codominant, or recessive patterns may be elucidated.

Continued advances in SNP maps and in high-throughput genotyping methods will facilitate the analysis of multiple polymorphisms within genes and the analysis of multiple genes within pathways. The effects of polymorphisms are best represented by their haplotypes. Data from multiple polymorphisms within a gene can be combined to create haplotypes, the set of multiple alleles on a single chromosome. None of the studies reviewed here reported haplotype associations, although several studies analyzed multiple polymorphisms within a gene, sometimes with inconsistent results. The analysis of haplotypes can increase the power to detect disease associations because of higher heterozygosity and tighter linkage disequilibrium with disease-causing mutations. In addition, haplotype analysis offers the advantage of not assuming that any of the genotyped polymorphisms is functional; rather, it allows for the possibility of an ungenotyped functional variant to be in linkage disequilibrium with the genotyped polymorphisms [67]. An analysis of data from multiple genes within the same DNA-repair pathway (particularly those known to form complexes) can provide more comprehensive insight into the studied associations. Such an analysis may shed light on the complexities of the many pathways involved in DNA repair and lung cancer development, providing hypotheses for future functional studies. Because of concerns over inflated type I error rates in pathway-wide or genome-wide association studies, methods of statistical analysis seeking to obviate this problem are under development [68]. The ability to include haplotype information and data from multiple genes, and to model their interactions, will provide more powerful and more comprehensive assessments of the DNA repair pathways.

This review, which is limited by the bias against publication of null findings, highlights the complexities inherent in epidemiological research and, particularly, in molecular epidemiological research. There is evidence that some polymorphisms in DNA repair genes play a role in carcinogenesis, most notably the *ERCC2/XPD* Lys751Gln and *XPA* G23A polymorphisms. The variant allele of each of the three polymorphisms was associated with about a 30% decrease or increase in lung cancer risk. Although the summary risk for developing lung cancer in individuals of each genotype may not be large, lung cancer is such a common malignancy that even a small increase in risk can translate to a large number of excess lung cancer cases. Therefore, polymorphisms, even those not strongly associated with lung cancer, should be considered as potentially important public health issues. In addition, it is important to keep in mind that a susceptibility factor in one population may not be a factor in another. There are differences in the prevalence of DNA repair polymorphisms across populations. In a population where the prevalence of an "at-risk" genotype in a given polymorphism is very low, the "at-risk" allele or "at-risk" genotype may be too infrequent to assess its associated risk. At a population level, the attributable risk must be small simply because it is an infrequent allele. Finally, the major bur-

den of lung cancer in the population probably results from the complex interaction between many genetic and environmental factors over time. Most environmental carcinogens first require metabolic activation by Phase I enzymes to their ultimate forms which then bind to DNA, forming aromatic-DNA adducts that are thought to be an early step in tumorigenesis. On the other hand, these activated forms are detoxified by Phase II enzymes. Thus, genetically determined susceptibility to lung cancer may depend on the metabolic balance among Phase I enzymes, Phase II enzymes and DNA repair enzymes [69]. Further investigations of the combined effects of polymorphisms between DNA repair genes and drug-metabolizing genes may also help to clarify the influence of genetic variation in the carcinogenic process. Consortia and international collaborative studies, which may be a way to maximize study efficacy and overcome the limitations of individual studies, are needed to help further illuminate the complex landscape of lung cancer risk and genetic variations.

Acknowledgements

Conflicts of interest

The authors have declared that no conflict of interest exists.

References

1. Pharoah PD, Dunning A, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 2004; 4: 850-60.
2. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes. Mutat Res 2005; 577: 275-83.
3. Cleaver JE. Defective repair replication of DNA in xeroderma pigmentosum. Nature 1968; 218: 652-6.
4. de Boer J, Hoeijmakers JH. Nucleotide excision repair and human syndromes. Carcinogenesis 2000; 21: 453-60.
5. Hanawalt PC. Subpathways of nucleotide excision repair and their regulation. Oncogene 2002; 21: 8949-56.
6. Bootsma D, Kraemer KH, Cleaver JE, Hoeijmakers JH. Nucleotide excision repair syndromes: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. In:Vogelstein B, Kinzler KW ed. The Genetic Basis of Human Cancer. New York: McGraw-Hill; 1998: 245-74.
7. Kraemer KH. Sunlight and skin cancer: another link revealed. Proc Natl Acad Sci (USA) 1997; 94: 11-4.
8. Lehmann AR. DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. Biochimie 2003; 85: 1101-11.
9. Park JY, Park SH, Choi JE, et al. Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2002; 11: 993-7.
10. Wu X, Zhao H, Wei Q, et al. XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. Carcinogenesis 2003; 24: 505-9.
11. Popanda O, Schattenberg T, Phong CT, et al. Specific combinations of DNA repair gene variants and increased risk for non-small cell lung cancer. Carcinogenesis 2004; 25: 2433-41.
12. Vogel U, Overvad K, Wallin H, et al. Combinations of polymorphisms in XPD, XPC and XPA in relation to risk of lung cancer. Cancer Lett 2005; 222: 67-74.
13. Zienolddiny S, Campa D, Lind H, et al. Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. Carcinogenesis 2006; 27: 560-7.
14. Vogel U, Laros I, Jacobsen NR, et al. Two regions in chromosome 19q13.2-3 are associated with risk of lung cancer. Mutat Res 2004; 546: 65-74.
15. Zhou W, Liu G., Park S, et al. Gene-smoking interaction associations for the ERCC1 polymorphisms in the risk of lung cancer. Cancer Epidemiol Biomarkers Prev 2005; 14: 491-6.
16. Yin J, Vogel U, Guo L, et al. Lack of association between DNA repair gene ERCC1 polymorphism and risk of lung cancer in a Chinese population. Cancer Genet Cytogenet 2006; 164: 66-70.
17. Matullo G., Dunning AM, Guarrera S, et al. DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. Carcinogenesis 2006; 27: 997-1007.
18. Zhou W, Liu G., Miller DP, et al. Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res 2002; 62: 1377-81.
19. Liang G., Xing D, Miao X, et al. Sequence variations in the DNA repair gene XPD and risk of lung cancer in a Chinese population. Int J Cancer 2003; 105: 669-73.
20. Butkiewicz D, Rusin M, Enewold L, et al. Genetic polymorphisms in DNA repair genes and risk of lung cancer. Carcinogenesis 2001; 22: 593-7.
21. Spitz MR, Wu X, Wang Y, et al. Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. Cancer Res 2001; 61: 1354-7.
22. Hou SM, Falt S, Angelini S, et al. The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis 2002; 23: 599-603.
23. Misra R., Ratnasinghe D, Tangrea JA, et al. Polymorphisms in the DNA repair genes XPD, XRCC1, XRCC3, and APE/ref-1, and the risk of lung cancer among male smokers in Finland. Cancer Lett 2003; 191: 171-8.
24. Shen M, Berndt SI, Rothman N, et al. Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. Int J Cancer 2005; 116: 768-73.
25. David-Beabes G.L, Lunn RM, London SJ. No association between the XPD (Lys751G1n) polymorphism or the XRCC3 (Thr241Met) polymorphism and lung cancer risk. Cancer Epidemiol Biomarkers Prev 2001; 10: 911-2.
26. Chen S, Tang D, Xue K, et al. DNA repair gene XRCC1 and XPD polymorphisms and risk of lung cancer in a Chinese population. Carcinogenesis 2002; 23: 1321-5.
27. Park JY, Lee SY, Jeon HS, et al. Lys751Gln polymorphism in the DNA repair gene XPD and risk of primary lung cancer. Lung Cancer 2002; 36: 15-6.
28. Harms C, Salama SA, Sierra-Torres CH, et al. Polymorphisms in DNA repair genes, chromosome aberrations, and lung cancer. Environ Mol Mutagen 2004; 44: 74-82.
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177-88.
30. Cochran WG. The combination of estimates from different experiments. Biometrics 1954; 10: 101-29.
31. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991; 10: 1665-7.
32. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994; 50: 1088-101.
33. Egger M, Davey SG., Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315: 629-34.
34. Sutton AJ, Duval SJ, Tweedie RL, et al. Empirical assessment of effect of publication bias on meta-analyses. BMJ 2000; 320: 1574-7.

35. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 1999; 91: 1194-210.

36. Livneh Z. DNA damage control by novel DNA polymerases: translesion replication and mutagenesis. J Biol Chem 2001; 276: 25639-42.

37. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J Natl Cancer Inst 2000; 92: 874-97.

38. Wei Q, Cheng L, Amos CI, et al. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. J Natl Cancer Inst 2000; 92: 1764-72.

39. Wei Q, Cheng L, Hong WK, Spitz MR. Reduced DNA repair capacity in lung cancer patients. Cancer Res 1996; 56: 4103-7.

40. Rajaee-Behbahani N, Schmezer P, Risch A, et al. Altered DNA repair capacity and bleomycin sensitivity as risk markers for non-small cell lung cancer. Int J Cancer 2001; 95: 86-91.

41. Li L, Lu X, Peterson CA, Legerski RJ. An interaction between the DNA repair factor XPA and replication protein A appears essential for nucleotide excision repair. Mol Cell Biol 1995; 15: 5396-402.

42. Sugasawa K, Ng JM, Masutani C, et al. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell 1998; 2: 223-32.

43. Reardon JT, Nichols AF, Keeney S, et al. Comparative analysis of binding of human damaged DNA-binding protein (XPE) and Escherichia coli damage recognition protein (UvrA) to the major ultraviolet photoproducts: T(c,s)T, T(t,s)T, T(6-4)T, and T(Dewar)T. J Biol Chem 1993; 268: 21301-8.

44. Wakasugim M, Shimizum M, Moriokam H, et al. Damaged DNA-binding protein DDB stimulates the excision of cyclobutane pyrimidine dimers in vitro in concert with XPA and replication protein A. J Biol Chem 2001; 276: 15434-40.

45. Butkiewicz D, Rusin M, Harris CC, Chorazy M. Identification of four single nucleotide polymorphisms in DNA repair genes: XPA and XPB (ERCC3) in Polish population. Hum Mutat 2000; 15: 577-8.

46. Kozak M. Interpreting cDNA sequences: some insights from studies on translation. Mamm Genome 1996; 7: 563-74.

47. Shen MR, Jones IM, Mohrenweiser H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. Cancer Res 1998; 58: 604-8.

48. Yu JJ, Mu C, Lee KB, et al. A nucleotide polymorphism in ERCC1 in human ovarian cancer cell lines and tumor tissues. Mutat Res 1997; 382: 13-20.

49. Coin F, Bergmann E, Tremeau-Bravard A, Egly JM. Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. EMBO J 1999; 18: 1357-66.

50. Hu Z, Wei Q, Wang X, Shen H. DNA repair gene XPD polymorphism and lung cancer risk: a meta-analysis. Lung Cancer 2004; 46: 1-10.

51. Benhamou S, Sarasin A. ERCC2 /XPD gene polymorphisms and lung cancer: a HuGE review. Am J Epidemiol 2005; 161: 1-14.

52. Modrich P, Lahue R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Ann Rev Biochem 1996; 65: 101-33.

53. Sijbers AM, de Laat WL, Ariza RR, et al. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell 1996; 86: 811-22.

54. Smith TR, Levine EA, Perrier ND, et al. DNA-repair genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2003; 12: 1200-4.

55. Fan F, Liu C, Tavare S, Arnheim N. Polymorphisms in the human DNA repair gene XPF. Mutat Res 1999; 406: 115-20.

56. Jeon HS, Kim KM, Park SH, et al. Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. Carcinogenesis 2003; 24: 1677-81.

57. Cui Y, Morgenstern H, Greenland S, et al. Polymorphism of Xeroderma Pigmentosum group G and the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus. Int J Cancer 2006; 118: 714-20.

58. Duell EJ, Wiencke JK, Chen TJ, et al. Polymorphisms in the DNA repair genes XRCC1 and ERCC2 and biomarkers of DNA damage in human blood mononuclear cells. Carcinogenesis 2000; 21: 965-71.

59. Matullo G., Peluso M, Polidoro S, et al. Combination of DNA repair gene single nucleotide polymorphisms and increased levels of DNA adducts in a population-based study. Cancer Epidemiol Biomarkers Prev 2003; 12: 674-7.

60. Palli D, Russo A, Masala G., et al. DNA adduct levels and DNA repair polymorphisms in traffic-exposed workers and a general population sample. Int J Cancer 2001; 94: 121-7.

61. Kozak M. Role of ATP in binding and migration of 40S ribosomal subunits. Cell 1990; 22: 459-67.

62. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene 1999; 234: 187-208.

63. Kozak M. Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. Proc Natl Acad Sci (USA) 1995; 92: 2662-6.

64. Tanaka K, Miura N, Satokata I, et al. Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. Nature 1990; 348: 73-6.

65. Sarasin A. An overview of the mechanisms of mutagenesis and carcinogenesis. Mutat Res 2003; 544: 99-106.

66. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 2004; 96: 434-42.

67. Khoury M, Beaty TH, Cohen BH. () Fundamentals of Genetic Epidemiology. Monographs in Epidemiology and Biostatistics. New York: Oxford University Press; 1993.

68. Hoh J, Wille A, Ott J. Trimming, weighting, and grouping SNPs in human case-control association studies. Genome Res 2001; 11: 2115-9.

69. Kiyohara C, Otsu A, Shirakawa T, et al. Genetic polymorphisms and lung cancer susceptibility: a review. Lung Cancer 2002; 37: 241-56.

Tables

Table 1. Genetic polymorphisms in the NER pathway and lung cancer risk: *XPA* G23A polymorphism

Author, published year (reference no.)	Ethnicity	No. of Cases /Controls	Source of controls	Frequency of A allele (p*)	OR (95% CI)**		Quality control of genotyping
					G/A	G/G	
Park et al., 2002 [9]	Asian	265/185	Population	0.478 (0.140)	1.00 (0.62 - 1.62)	0.62 (0.35 - 1.10)	Sequencing
Wu et al., 2003 [10]	Caucasian	564/581	Population	0.446 (0.066)	0.65 (0.48 - 0.87)	0.74 (0.55 - 1.01)	None
Wu et al., 2003 [10]	Mexican-American	50/47	Population	0.394 (0.057)	0.31 (0.09 - 1.00)	0.40 (0.13 - 1.25)	None
Wu et al., 2003 [10]	African-American	71/67	Population	0.299 (0.193)	0.54 (0.16 - 1.68)	0.49 (0.15 - 1.49)	None
Popanda et al., 2004 [11]	Caucasian	461/457	Hospital	0.334 (0.682)	0.77 (0.48 - 1.21)	0.82 (0.52 - 1.30)	Replication (random samples)
Vogel et al., 2005 [12]	Caucasian	256/269	Population	0.268 (0.019)	0.78 (0.41 - 1.49)	0.57 (0.30 - 1.06)	None
Zienolddiny et al., 2006 [13]	Caucasian	248/276	Population	0.361 (0.033)	0.87 (0.48 - 1.57)	1.41 (0.79 - 2.52)	Replication (all samples)
Summary†	No. of populations						
All	7	1913/1882		0.368 (0.308 - 0.429)	0.73 (0.61 - 0.89) p‡ = 0.562	0.75 (0.59 - 0.95) p‡ = 0.272	
Caucasian	4	1527/1583		0.352 (0.277 - 0.428)	0.72 (0.58 - 0.89) p‡ = 0.805	0.82 (0.61 - 1.11) p‡ = 0.169	

* P for difference of allelic frequency between cases and controls.

** Crude odds ratio and 95% confidence interval.

† Based on random effects model.

‡ P for heterogeneity (Cochran Q test).

Table 2. Genetic polymorphisms in the NER pathway and lung cancer risk: *ERCC1* T19007C polymorphism

Author, published year (reference no.)	Ethnicity	No. of Cases /Controls	Source of controls	Frequency of T allele (p*)	OR (95% CI)**		Quality control of genotyping
					T/C	C/C	
Vogel et al., 2004 [14]	Caucasian	252/266	Population	0.617 (0.632)	0.99 (0.67 - 1.47)	0.86 (0.49 - 1.50)	Replication (random samples)
Zhou et al., 2005 [15]	Caucasian	1752/1358	Population	0.609 (0.875)	1.00 (0.86 - 1.17)	1.02 (0.82 - 1.27)	Replication (random samples)
Yin et al., 2006 [16]	Asian	151/143	Hospital	0.203 (0.940)	0.39 (0.08 - 1.53)	0.49 (0.11 - 1.84)	Replication (random samples)
Zienolddiny et al., 2006 [13]	Caucasian	260/213	Population	0.462 (00004)	0.50 (0.31 - 0.79)	0.35 (0.20 - 0.61)	Replication (all samples)
Matullo et al., 2006 [17]	Caucasian	116/1093	Population	0.598 (0.423)	0.82 (0.52 - 1.27)	0.85 (0.46 - 1.51)	Replication (random samples)
Summary†	No. of populations						
All	5	2531/3073		0.499 (0.387 - 0.611)	0.82 (0.62 - 1.08) p‡ = 0.053	0.72 (0.46 - 1.11) p‡ = 0.012	
Caucasian	4	2380/2930		0.575 (0.529 - 0.622)	0.84 (0.63 - 1.11) p‡ = 0.046	0.74 (0.46 - 1.17) p‡ = 0.007	

* P for difference of allelic frequency between cases and controls.

** Crude odds ratio and 95% confidence interval.

† Based on random effects model.

‡ P for heterogeneity (Cochran Q test).

Table 3. Genetic polymorphisms in the NER pathway and lung cancer risk: *ERCC2* Asp312Asn polymorphism

Author, published year (reference no.)	Ethnicity	No. of Cases /Controls	Source of controls	Frequency of Asp allele (p*)	OR (95% CI)**		Quality control of genotyping
					Asp/Asn	Asn/Asn	
Butkiewicz et al., 2001 [20]	Caucasian	96/94	Population	0.564 (0.187)	0.49 (0.24 - 0.98)	0.71 (0.29 - 1.74)	Sequencing
Spitz et al., 2001 [21]	Admixed population	195/257	Population	0.728 (0.509)	0.92 (0.62 - 1.36)	1.54 (0.78 - 3.05)	None
Hou et al., 2002 [22]	Caucasian	184/162	Population	0.630 (0.900)	1.27 (0.78 - 2.05)	0.88 (0.43 - 1.84)	Replication (random samples)
Zhou et al., 2002 [18]	Caucasian	1092/1240	Population	0.669 (0.498)	0.98 (0.82 - 1.17)	1.41 (1.06 - 1.86)	Replication (random samples)
Liang et al., 2003 [23]	Asian	1006/1020	Population	0.935 (0.294)	0.98 (0.76 - 1.28)	11.2 (1.45 - 87.2)	Replication (random samples)
Misra et al., 2003 [24]	Caucasian	313/312	Population	0.636 (0.384)	0.76 (0.53 - 1.07)	0.94 (0.56 - 1.59)	Replication (random samples)
Popanda et al., 2004 [11]	Caucasian	463/460	Hospital	0.630 (0.674)	1.14 (0.77 - 1.68)	1.03 (0.70 - 1.51)	Replication (random samples)
Vogel et al., 2004 [14]	Caucasian	252/263	Population	0.644 (0.475)	1.27 (0.86 - 1.89)	1.09 (0.63 - 1.86)	None
Shen et al., 2005 [25]	Asian	118/113	Population	0.938 (0.239)	0.58 (0.21 - 1.52)	—	Replication (random samples)
Zienolddiny et al., 2006 [13]	Caucasian	275/290	Population	0.622 (0.884)	0.85 (0.58 - 1.25)	1.11 (0.68 - 1.81)	Replication (all samples)
Matullo et al., 2006 [17]	Caucasian	116/1094	Population	0.613 (0.635)	0.81 (0.52 - 1.26)	0.95 (0.51 - 1.71)	Replication (random samples)
Summary†	No. of populations						
All	11	4110/5305		0.692 (0.591 - 0.794)	0.95 (0.84 - 1.07) p‡ = 0.342	1.14 (0.95 - 1.37) p‡ = 0.317	
Caucasian	8	2791/3915		0.645 (0.572 - 0.719)	1.12 (0.95 - 1.32) p‡ = 0.178	1.12 (0.95 - 1.32) p‡ = 0.672	
Asian	2 (1)	1124/1133		0.936 (0.925 - 0.946)	0.95 (0.73 - 1.23) p‡ = 0.315	—	

* P for difference of allelic frequency between cases and controls.

** Crude odds ratio and 95% confidence interval.

† Based on random effects model.

‡ P for heterogeneity (Cochran Q test).

Table 4. Genetic polymorphisms in the NER pathway and lung cancer risk: *ERCC2* Lys751Gln polymorphism

Author, published year (reference no.)	Ethnicity	No. of Cases /Controls	Source of controls	Frequency of Lys allele (p*)	OR (95% CI)**		Quality control of genotyping
					Lys/Gln	Gln/Gln	
David-Beabes et al., 2001 [25]	Caucasian	178/453	Population	0.653 (0.044)	1.14 (0.77 - 1.71)	1.72 (1.00 - 2.94)	Replication (random samples)
David-Beabes et al., 2001 [25]	African-American	153/234	Population	0.750 (0.390)	1.14 (0.73 - 1.78)	1.39 (0.54 - 3.55)	Replication (random samples)
Spitz et al., 2001 [21]	Admixed population	341/360	Population	0.679 (0.257)	1.07 (0.78 - 1.46)	1.36 (0.84 - 2.20)	None
Chen et al., 2002 [26]	Asian	109/109	Population	0.596 (0.050)	0.79 (0.17 - 1.11)	0.44 (0.17 - 1.11)	None

Author, published year (reference no.)	Ethnicity	No. of Cases /Controls	Source of controls	Frequency of Lys allele (p*)	OR (95% CI)**		Quality control of genotyping
					Lys/Gln	Gln/Gln	
Hou et al., 2002 [22]	Caucasian	185/162	Population	0.627 (0.568)	1.22 (0.75 - 2.00)	1.11 (0.58 - 2.13)	Replication (random samples)
Zhou et al., 2002 [18]	Caucasian	1092/1240	Population	0.634 (0.313)	1.01 (0.84 - 1.21)	1.17 (0.90 - 1.51)	Replication (random samples)
Park et al., 2002 [27]	Asian	250/163	Population	0.945 (0.687)	1.06 (0.55 - 2.11)	—	None
Liang et al., 2003 [19]	Asian	1006/1020	Population	0.913 (0.762)	0.93 (0.73 - 1.18)	2.36 (0.90 - 6.17)	Replication (random samples)
Misra et al., 2003 [23]	Caucasian	310/302	Population	0.594 (0.978)	0.87 (0.60 - 1.26)	1.06 (0.64 - 1.76)	Replication (random samples)
Popanda et al., 2004 [11]	Caucasian	463/459	Hospital	0.635 (0.104)	1.14 (0.86 - 1.52)	1.37 (0.93 - 2.02)	Replication (random samples)
Harms et al., 2004 [28]	Caucasian	110/119	Population	0.727 (0.458)	1.34 (0.79 - 2.49)	1.07 (0.34 - 3.38)	Replication (all samples)
Vogel et al., 2004 [14]	Caucasian	256/269	Population	0.652 (0.009)	1.57 (1.05 - 2.34)	1.73 (1.01 - 2.96)	None
Shen et al., 2005 [24]	Asian	118/108	Population	0.889 (0.010)	0.44 (0.18 - 1.03)	—	Replication (random samples)
Zienolddiny et al., 2006 [13]	Caucasian	317/386	Population	0.631 (0.007)	1.20 (0.84 - 1.73)	1.56 (1.06 - 2.31)	Replication (all samples)
Matullo et al., 2006 [17]	Caucasian	116/1094	Population	0.594 (0.475)	1.23 (0.78 - 1.96)	1.17 (0.63 - 2.11)	Replication (random samples)
Summary†	No. of populations						
All	15 (13)	5004/6478		0.701 (0.622 - 0.779)	1.06 (0.97 - 1.16) p‡ = 0.505	1.30 (1.13 - 1.49) p‡ = 0.495	
Caucasian	9	3027/4484		0.634 (0.614 - 0.655)	1.11 (1.00 - 1.24) p‡ = 0.587	2.25 (0.97 - 5.23) p‡ = 0.785	
Asian	4 (2)	1484/1400		0.843 (0.763 - 0.924)	0.89 (0.72 - 1.09) p‡ = 0.386	1.02 (0.20 - 5.27) p‡ = 0.014	

* P for difference of allelic frequency between cases and controls.

** Crude odds ratio and 95% confidence interval.

‡ P for heterogeneity (Cochran Q test).