Carcinogenesis vol.28 no.3 pp.507–518, 2007
doi:10.1093/carcin/bgl253
Advance Access publication December 20, 2006

REVIEW

The molecular epidemiology of lung cancer

Ann G.Schwartz*, Geoffrey M.Prysak, Cathryn H.Bock and Michele L.Cote

Population Studies and Prevention Program, Karmanos Cancer Institute, and Department of Internal Medicine, Wayne State University School of Medicine, 110 E. Warren Avenue, Detroit, MI 48103, USA

*To whom correspondence should be addressed. Tel: +313 833 0715 ext. 2024; Fax: +313 833 7806;
Email: schwarta@karmanos.org

Lung cancer is the leading cause of cancer mortality worldwide. There have been only slight improvements in early diagnosis and survival, reflecting limited advances in screening and treatment for lung cancer. The identification of host differences in susceptibility to lung carcinogens, in particular to cigarette smoke, is essential in predicting who is at highest risk. Susceptibility differences in the form of rare, high-penetrance genes are suggested from studies of familial aggregation of lung cancer and a linkage study. Studies focused on more common, low-penetrance genes in the tobacco smoke metabolism pathways (phase I and phase II enzymes) and DNA repair pathways are reviewed, as are inflammation and cell cycle-related genes and DNA adducts as intermediate biomarkers. Also reviewed are studies of epigenetic mechanisms, such as methylation, as alternative sources of variation in host susceptibility. Studies of molecular epidemiology in lung cancer survival are discussed briefly. In the future, studies that focus on complex interactions between multiple genes and environmental exposures within pertinent pathways are needed. New technological advances in genotyping will help move the field forward.

Introduction

Lung cancer is the leading cause of cancer mortality worldwide. In the USA in 2006, an estimated 174 470 new lung cancer diagnoses and 162 460 deaths will occur (1). Whereas men have seen declines in incidence and mortality beginning in the mid-1990s, in women, incidence rates have only recently leveled off and mortality rates continue to increase (2). Survival after a lung cancer diagnosis has changed little, with overall 5-year relative survival increasing only slightly from 12.4% for 1974–1976 diagnoses to 15.0% for 1996–2002 diagnoses (1). This slight improvement in survival reflects limited advances in screening and treatment for lung cancer.

Although cancers of the lung have proven difficult to diagnose early and treat successfully, the cause of most of these cancers is well known. Epidemiologic studies have demonstrated that cigarette smoking is the major risk factor in the development of lung cancer (3), with a striking dose-response relationship (4). It is estimated that 80–90% of lung cancer incidence can be attributed to cigarette smoking (5). Even with the high attributable risks due to cigarette smoke exposure, only 10–15% of all smokers develop lung cancer (6), suggesting that there are host differences in susceptibility to lung carcinogens. Predicting which smokers are at highest risk would focus screening and chemoprevention studies and offer insights into biologic mechanisms.

Abbreviations: ADPRT, adenosine diphosphate ribosyl transferase; BER, base excision repair; CI, confidence interval; COX2, cyclooxygenase-2; GSEC, Genetic Susceptibility to Environmental Carcinogenesis; GST, glutathione *S*-transferase; HR, hazard ratio; IL, interleukin; MDM2, mouse double minute 2; mEH, microsomal epoxide hydrolase; MMPs, matrix metalloproteinases; MPO, myeloperoxidase; NER, nucleotide excision repair; NQO1, NAD(P)H quinone oxidoreductase 1; NSCLC, non-small-cell lung cancer; OR, odds ratio; RR, relative risk; SNP, single-nucleotide polymorphism; SOD, superoxide dismutase.

Technologic advances over the last 20 years have allowed for the investigation of the molecular mechanisms underlying susceptibility to lung cancer and provided the necessary tools for molecular epidemiology studies. Susceptibility differences may be inherited in the form of low-frequency, high-penetrance genes or high-frequency, low-penetrance genes or may be acquired through epigenetic mechanisms such as methylation, with likely genetic heterogeneity. The goal of this review is to summarize the molecular epidemiology of lung cancer, focusing on larger studies (including pooled and meta-analyses) and new areas of research on inherited and acquired susceptibility and complex interactions between multiple genes and environmental exposures within pertinent pathways. Studies of the molecular epidemiology of lung cancer survival are discussed briefly.

High-penetrance, low-frequency genes

Many lines of evidence support a hereditary influence on risk of lung and other smoking-related cancers. Familial aggregation is a hallmark for inherited susceptibility, particularly when familial clustering of shared environmental factors, such as smoking, can be ruled out. Over the last 40 years, studies with increasingly sophisticated designs and analytic approaches have consistently demonstrated familial aggregation, even after adjusting for family smoking patterns (7–12). A recent meta-analysis including 32 studies revealed an ∼2-fold increased risk of lung cancer in individuals with a family history [case–control studies: odds ratio (OR) = 1.82, 95% confidence interval (CI) 1.58–2.10; cohort studies: Relative risk = 2.01, 95% CI 1.62–2.50] (8). In the 11 studies that evaluated risk among non-smokers, risk associated with family history was still elevated (RR = 1.51, 95% CI 1.11–2.06). It is well recognized that there are limitations to these studies, the most important of which are incomplete adjustment for smoking among relatives and lack of validation of family histories. However, even with these limitations, the consistency of findings across study designs and populations lends credence to these results.

Family linkage studies have been used successfully to identify highly penetrant, low-frequency susceptibility genes for other cancers, such as *BRCA1* and *BRCA2* for breast cancer and *APC* for colorectal cancer. Bailey-Wilson *et al.* (13) report the first results of a region on chromosome 6q23–25 (146cM–164cM) linked to familial lung cancer. Multipoint parametric linkage analysis was conducted assuming an autosomal dominant, low-penetrance model with a susceptibility allele frequency of 0.01, 10% penetrance in gene carriers and 1% penetrance in non-carriers. Under this model, a maximum HLOD (Logarithm of the Odds favoring genetic linkage given Heterogeneity) of 2.79 on chromosome 6q (155cM) was achieved. In a subset of 23 families with five or more affected relatives in multiple generations, the HLOD was 4.26, with 94% of these families estimated to be linked to this region. The 1-HLOD support interval in these 23 families extended from 146cM to 164cM. Non-parametric tests, which do not require specification of an underlying inheritance model, supported the parametric linkage findings. On top of the effect of inheritance, smoking was associated with a 3-fold increase in lung cancer risk. The search for a lung cancer gene within this region is ongoing. Only ∼1% of lung cancer patients have such extensive family histories (with three or more affected relatives), but 10–15% have at least one first-degree relative with the disease and host susceptibility differences exist regardless of family history, so the search for susceptibility genes has also focused on genes with minor allele frequencies of >5%.

Low-penetrance, high-frequency genes

Candidate susceptibility genes for lung cancer have been extensively studied, with most of the work focusing on mechanistically plausible

variants in genes coding for enzymes involved in the activation, detoxification and repair of damage caused by tobacco smoke. Alterations in these pathways are hypothesized to affect an individual's processing of tobacco carcinogens and therefore risk of developing lung cancer. In addition, genes in inflammatory pathways have been studied.

Carcinogen-metabolizing genes

Some of the most widely studied polymorphic loci are those coding for phase I and II enzymes involved in the activation and conjugation of tobacco smoke. Metabolism of polycyclic aromatic hydrocarbons (PAHs), tobacco-specific nitrosamines and aromatic amines in cigarette smoke occurs via two classes of enzymes: phase I (oxidation/reduction/hydrolysis) and phase II (conjugation) enzymes. The most frequently studied phase I and II enzymes include CYP1A1, microsomal epoxide hydrolase 1 (mEH/EPHX1), myeloperoxidase (MPO), NAD(P)H quinone oxidoreductase 1 (NQO1) and the glutathione *S*-transferases (GST: M1, P1 and T1), although others have also been studied. A survey of PubMed (English language, original journal articles) for the past 10 years uncovered 92 studies of *CYP1A1* variants with an average sample size of 120 cases and 200 controls, 103 studies of *GSTM1* variants with an average sample size of ~230 cases and 315 controls and between 10 and 20 studies were identified for *CYP1B1*, *CYP2D6*, *CYP2E1*, *NQO1*, *EPHX1* and *MPO*. Results from pooled and meta-analyses are presented in Table I.

CYP1A1 is active in the metabolism of PAHs found in tobacco smoke. Evidence for an association between *CYP1A1* polymorphisms and risk of lung cancer originally came from studies in Japanese populations where variant alleles occur at higher frequencies than in Caucasians, with reports of >2-fold increased risk (14). Other studies have reported conflicting results (15–20). A meta-analysis of 16 studies of *CYP1A1* polymorphisms found no association with lung cancer risk (MspI: OR = 1.27, 95% CI 0.91–1.77; Ile462Val Val/Val homozygotes: OR = 1.62, 95% CI 0.93–2.82) (21). Both a pooled analysis and a meta-analysis of the *CYP1A1* Ile462Val polymorphism were conducted by the Genetic Susceptibility to Environmental Carcinogenesis (GSEC) consortium (22). The meta-analysis demonstrated no significant risk associated with the presence of the Val allele, whereas the pooled analysis found a significant 55% increased risk in Caucasians heterozygous at this locus. The effect of the polymorphism was stronger for females than for males, and in never-smokers than in ever-smokers. In another pooled analysis using data from 22 studies, a significant 2.4-fold increase in risk was seen in individuals carrying the MspI variant (23). In the pooled data, those under the age of 45 demonstrated a 4.7-fold increased risk, but this finding was based on very small numbers (five cases and four controls) (24). A pooled analysis of 14 studies in non-smokers by GSEC showed a 3-fold increased risk associated with the Val allele of Ile462Val (95% CI 1.51–5.91), again based on small numbers, but no effect of the MspI polymorphism (25). Findings in non-smokers or light smokers are not uncommon in genetic epidemiology studies of lung cancer and have been interpreted to suggest that in heavy smokers the exposure is so high that subtle changes in enzyme function based on genotype are masked. Non-smokers have lower exposures to lung carcinogens in tobacco smoke (primarily through environmental tobacco smoke exposure) so contributions from genetic polymorphisms may be stronger in this subpopulation. The evaluation of risk in non-smokers unfortunately includes small numbers of individuals. Only one study was identified that evaluated *CYP1A1* variants in survival after a lung cancer diagnosis. In this relatively small study of 232 patients, those carrying at least one copy of the MspI variant showed decreased survival (26).

GSTs occur in a number of classes and act to conjugate electrophilic compounds with glutathione. *GSTM1* occurs in the null form in

Table I. Risk associated with genetic polymorphisms in phase I and phase II carcinogen-metabolizing enzymes

Author	No. of studies (cases/controls)	OR	95% CI	Risk genotype	Comments
CYP1A1 MspI and *CYP1A1* Ile462Val					
Houlston[a] (21)	16 (total 3473)	1.27	0.91–1.77	MspI homozygous variant	No smoking or race data
	8 (total 2290)	1.62	0.93–2.82	Ile462Val Val/Val	No smoking or race data
Vineis *et al.*[b] (23)	22 (2451/3358)	2.36	1.16–4.81	MspI homozygous variant	Caucasians, 30 cases/24 controls with risk genotype
Taioli *et al.*[b] (24)	21 (261/1452)	4.7	1.2–19.0	MspI homozygous variant	Caucasians, ≤45 years of age, five cases/four controls with risk genotype
	21 (261/1452)	3.3	0.2–44.5	Ile462Val Val/Val	Caucasians, ≤45 years of age, one case/two controls with risk genotype
Le Marchand *et al.* (22)	11 (1950/2617)	1.28[a]	0.94–1.75	Ile462Val Val allele	Caucasians
	11 (1950/2617)	1.55[b]	1.11–2.16	Ile462Val Val allele	Caucasians
Hung *et al.*[b] (25)	14 (302/1631)	2.99	1.51–5.91	Ile462Val Val allele	Non-smokers
	14 (302/1631)	2.17	1.12–4.12	MspI variant allele	Non-smokers
GSTM1					
Houlston[a] (27)	21 (3593/6131)	1.13	1.04–1.25	Null	Only two studies had >80% power to detect risk of 1.5
Taioli *et al.*[b] (24)	22 (236/1452)	1.1	0.9–1.3	Null	Caucasians ≤45 years of age
Benhamou *et al.* (147)	43 (7463/10 789)	1.17[a]	1.07–1.27	Null	Lack of homogeneity across all studies
	21 (3940/5515)	1.08[b]	0.98–1.18	Null	No differences by smoking status, sex, histology and ethnicity
Ye *et al.*[a] (29)	130 (23 452/30 397)	1.18	1.14–1.23	Null	
	5 (3436/3897)	1.04	0.95–1.14	Null	Larger studies only (>500 cases)
Hung *et al.*[b] (25)	14 (302/1631)	1.15	0.86–1.53	Null	Non-smokers
MPO (G-463A)					
Feyler *et al.*[a] (32)	10 (2686/3325)	0.86	0.31–2.32	AA (versus GG)	Results influenced by a single large study
Kiyohara *et al.*[a] (36)	12 (4285/4656)	0.81	0.64–1.02	AA + GA (versus GG)	Multiple ethnic groups
NQO1 Pro187Ser					
Kiyohara *et al.*[a] (36)	3 (741/846)	1.12	0.96–1.47	Pro/Ser + Ser/Ser	Caucasians
	3 (499/959)	0.70	0.56–0.88	Pro/Ser + Ser/Ser	Japanese
mEH Tyr113His (decreased activity) and His138Arg (increased activity)					
Lee *et al.* (44)	7 (2078/3081)[a]	1.54	0.77–3.07	High-activity alleles	
	8 (986/1633)[b]	1.18	0.92–1.52	High-activity alleles	
Kiyohara *et al.*[b] (50)	8 (1944/2670)	0.75	0.53–1.07	Low-activity alleles	Nine ethnic groups
	6 (815/1286)	0.72	0.43–1.22	Low-activity alleles	Caucasians

[a]Meta-analyses.
[b]Pooled analyses.

~50% of the Caucasian population. One of the first meta-analyses conducted of *GSTM1*-null variants (27) showed a modest increase in lung cancer among carriers of the *GSTM1*-null genotype (OR = 1.13, 95% CI 1.04–1.25). A recent pooled analysis in Caucasians <45 years of age also found similar, but not statistically significant, results (OR = 1.1, 95% CI 0.9–1.3) (24). A larger meta-analysis reported that lung cancer risk increased by 17% in those who were *GSTM1* null (95% CI 1.07–1.27) (28). The companion pooled analysis based on 21 studies reported no significant findings. The most recent and largest meta-analysis of 130 studies found an 18% increased risk of lung cancer among individuals with the *GSTM1*-null genotype (95% CI 1.14–1.23), but when analyzing data only from the larger studies there was no association (29). This same group also looked at *GSTP1* and *GSTT1* in these 130 studies. For *GSTP1* I105V and A114V, there was no association with lung cancer risk, whereas for the *GSTT1*-null genotype, risk of lung cancer was modestly increased (OR = 1.09, 95% CI 1.02–1.16). However, when only the larger studies with a lower probability of false-positive findings were evaluated, the *GSTT1* association became non-significant. Both study size and ethnic background were important sources of heterogeneity between studies. Frequencies of variant alleles differ markedly by ethnicity and studies in populations with lower frequencies of variant alleles may be underpowered to detect risk differences. In addition to conjugating carcinogens in tobacco smoke, the GSTs catalyze glutathione conjugation of chemotherapeutic agents and by-products of reactive oxygen damage, suggesting that genotype might predict survival after a lung cancer diagnosis. Studies to date have been relatively small. Two of the largest studies have shown (i) increased risk of death for men carrying the *GSTM1*-null genotype, but no associations for *GSTT1* or *GSTP1* (30) and (ii) decreased risk of death for late-stage cases carrying the exon 6, and not the exon 5, variant of *GSTP1* (31). These studies are limited in that complete treatment data, including chemotherapy use and other measures of prognosis such as comorbidities, were not available.

MPO oxidizes pro-carcinogens and participates in locally mediated immune response. Individual studies investigating *MPO* polymorphisms have been small but have shown fairly consistent, although not always statistically significant, reductions in risk of lung cancer associated with the variant allele of ~50% (32–34). The largest study conducted, however, reported a non-significant OR of 1.15 in Caucasians (35). Ten case–control studies were combined to investigate the role of the *MPO* G-463A promoter polymorphism in lung cancer (32). No association between the AA genotype and lung cancer risk was found when analyzing all studies together. Twelve case–control studies were included in a meta-analysis that reported a non-significant summary OR for the AA genotype of 0.81 (36).

NQO1 can act in both carcinogen activation and detoxification. Studies that have reported significant associations between *NQO1* genotype and lung cancer risk have been based on fairly small sample sizes and/or subgroup analysis (37–40). In the largest study (40), risk was increased at lower levels of cigarette consumption and decreased at higher levels of cigarette consumption, suggesting complex gene–environment interactions. A recent meta-analysis reported that the Pro/Ser + Ser/Ser genotypes were not associated with risk in Caucasians (OR = 1.12, 95% CI 0.96–1.47) but were associated with reduced risk in Japanese (OR = 0.70, 95% CI 0.56–0.88) (36). Overall, however, there was significant heterogeneity between studies.

mEH has two roles. It hydrolyzes arenes, alkenes and aliphatic epoxides, making them less reactive and activates PAHs found in tobacco smoke into more reactive compounds. Two polymorphisms have been studied in the gene for mEH (*EPHX1*). The Tyr113His polymorphism is associated with decreased enzymatic activity and His139Arg is associated with increased activity. Studies in lung cancer have generally reported that the predicted high-activity genotype is associated with increased risk (36,41–48). One of the largest studies (974 cases and 1142 controls) demonstrated a 2-fold increased risk among non-smokers with the low-activity genotype (95% CI 1.1–3.3) and 30% decreased risk in smokers of 80 pack-years with the low-activity genotype (95% CI 0.4–1.0) (49). Given mEH's role in detoxification and activation, it is not surprising to see variation in results based on cigarette smoke exposure. A pooled analysis demonstrated no significant increased risk associated with these polymorphisms (44). A second meta-analysis that included both polymorphisms reported a non-significant summary OR of 0.75 for low-activity genotypes (50).

Gene–gene combinations/interactions. The results of large studies, meta-analyses and pooled analyses of candidate gene polymorphisms do not point conclusively to a single polymorphism in one gene that substantially alters the risk of lung cancer. The lack of significant findings that are reproducible across studies has refocused the thinking behind these studies and led to studies of multiple genes within pathways. To accomplish this, larger studies are needed and consortia have emerged to conduct pooled and meta-analyses. There have been many attempts to look at multiple genes in a small study setting, but this review is limited to larger studies (Table II).

The role of combined genotypes at *CYP1A1* MspI and Ile462Val, *GSTM1* null, *GSTT1* null and *NQO1* Pro187Ser was evaluated in a Swedish population (51). All possible combinations of four genes were explored with no evidence of gene–gene interactions found; however, the protective effect of the *GSTT1*-null genotype for squamous cell carcinoma of the lung reported was limited to those individuals carrying the *CYP1A1*1/*1* genotype (lacking both variants) (OR = 0.33, 95% CI 0.10–0.95).

Vineis *et al.* (52) report an analysis of multiplicative interactions between phase I and II metabolic enzymes, *CYP1A1*2A* (MspI), *GSTM1* null and *GSTT1* null, utilizing the GSEC collaborative consortium. This study was powered to detect ORs for interactions of ~1.5–3.0 for the various genotype combinations. Individual genotype analysis showed increased lung cancer risk associated with the *CYP1A1*2A* variant (OR = 2.6, 95% CI 1.2–5.7); however, no statistically significant interactions were evident.

In the work by Raimondi *et al.* (53), potential interactions between the phase I and II enzymes *CYP1A1* MspI, *CYP1A1* Ile462Val, *GSTM1* and *GSTT1* were evaluated in non-smokers from GSEC. Analyses were performed separately for the Caucasians and Asians to minimize bias from a heterogeneous population with different underlying allele frequencies. A protective effect was found for Caucasians with the combination genotype of *CYP1A1* non-variants, *GSTM1* null and *GSTT1* non-null (for *CYP1A1* MspI non-variant/*GSTM1* null/*GSTT1* non-null: OR = 0.34, 95% CI 0.15–0.76) (for *CYP1A1* Ile462Val Ile/Ile/*GSTM1* null/*GSTT1* non-null: OR = 0.29, 95% CI 0.13–0.62). Even beginning with a large study population, the number of individuals carrying one of these protective genotypes was 28 (5.1%) cases for the combination including the *CYP1A1* MspI polymorphism and 29 (5.2%) cases for the combination including the *CYP1A1* Ile462Val.

Larsen *et al.* (54) reported that individuals carrying two risk alleles in the carcinogen-metabolizing pathway (*MPO* G-463A GG and *CYP1A1* Ile462Val Ile/Val + Val/Val) were at increased risk of lung cancer in general (OR = 2.88, 95% CI 1.70–5.00) and adenocarcinomas in particular (OR = 3.72, 95% CI 2.01–6.88) versus those carrying the referent alleles (*MPO* AA/AG and *CYP1A1* Ile/Ile). This was the only significant interaction out of 10 possible two-gene combinations for *GSTM1*, *GSTP1*, *GSTT1*, *MPO* and *CYP1A1*. This study included 1103 cases and 627 chronic obstructive pulmonary disease and healthy smoking controls. The use of smoking controls may not only result in a conservative estimate of risk but also limit discovery of susceptibility genes if gene associations are only evident in those with low exposures.

Gene–gene interactions between *MPO* and manganese superoxide dismutase (*SOD2*) have also been explored in a study by Liu *et al.* (55). These enzymes are involved both in the generation and in the removal of reactive oxygen species, a by-product of inflammation associated with airway exposure to tobacco smoke, in addition to MPO's phase I metabolizing activity. In the subset of females carrying the *MPO* G-463A variant genotype (GA or AA), increased risk of lung cancer was associated with the *SOD2* Ala16Val risk genotype (Val/Val

Table II. Risks associated with multiple risk genotypes

Author	Cases/controls	OR	95% CI	Risk genotype	Comments
Phase I and phase I enzyme polymorphisms					
Alexandrie *et al.* (51)	524/530	0.33	0.10–0.95	*GSTT1* null and *CYP1A1* non-variants	Squamous cell only. No gene gene interactions found when evaluating all possible combinations of *CYP1A1* MspI, *CYP1A1* Ile462Val, *GSTM1*, *GSTT1* and *NQO1*
Vineis *et al.*[a] (52)	1361/1247	2.70	0.50–15.30	*CYP1A1*2A/GSTM1* null	Caucasians
		0.10	0.04–2.70	*CYP1A1*2A/GSTT1* null	Caucasians
		1.00	0.60–1.50	*GSTM1* null/*GSTT1* null	Caucasians
Raimondi *et al.*[a] (53)	555/2209	0.34	0.15–0.76	*CYP1A1* MspI non-variant/*GSTM1* null/*GSTT1* non-null	Caucasian, non-smokers
	555/2209	0.29	0.13–0.62	*CYP1A1* Ile/Ile/*GSTM1* null/*GSTT1* non-null	Caucasian, non-smokers
Larsen *et al.* (54)	1103/627	2.88	1.70–5.00	*MPO* GG/*CYP1A1* Val allele	Non-small-cell lung cancer
		3.72	2.10–6.88	*MPO* GG/*CYP1A1* Val allele	Adenocarcinoma
Liu *et al.* (55)	830/1119	3.26	1.55–6.83	*MPO* A allele/*SOD2* Val/Val	Females
		1.37	0.66–2.86	*MPO* A allele/*SOD2* Val/Val	Males
Zhou *et al.* (56)	1115/1250	0.47	0.27–0.83	*NAT2* slow/*mEH* high	Non-smokers
		0.30	0.14–0.62	*NAT2* rapid/*mEH* high	Non-smokers
		1.65	1.07–2.56	*NAT2* slow/*mEH* high	80 pack-year smokers
		2.19	1.26–3.81	*NAT2* rapid/*mEH* high	80 pack-year smokers
		2.19	1.22–3.95	*NAT2* slow/*mEH* high	120 pack-year smokers
		3.42	1.61–7.26	*NAT2* rapid/*mEH* high	120 pack-year smokers
DNA repair polymorphisms					
Zhang *et al.* (71)	1000/1000	1.66	1.09–1.63	*ADPRT* Ala/Ala/*XRCC1* Arg/Arg	Chinese, not adjusted for smoking
		5.91	2.09–16.72	*ADPRT* Ala/Ala/*XRCC1* Gln/Gln	Chinese, not adjusted for smoking
Zhou *et al.* (63)	1091/1240	5.2	1.7–16.6	Five or six risk alleles: *XRCC1* Arg399Gln, *XDP* Asp312Asn and Lys751Gln	Caucasians, non-smokers
	1091/1240	0.3	0.1–0.8	5 or 6 risk alleles: *XRCC1* Arg399Gln, *XDP* Asp312Asn and Lys751Gln	Caucasians, heavy smokers
Cell cycle polymorphisms					
Zhang *et al.* (92)	1106/1420	4.56	2.76–7.54	*MDM2* G-5 + 309T GG and *TP53* Arg72Pro Pro/Pro	Chinese
		10.41	5.26–20.6	*MDM2* G-5 + 309T GG and *TP53* Arg72Pro Pro/Pro	Chinese, smokers
Schabath *et al.* (93)	863/852	1.30	1.05–1.61	1–3 *p53* + *p73* risk alleles	Hospital based
		1.778	1.23–2.56	≥4 *p53* + *p73* risk alleles	Hospital based
Zhang *et al.* (92)	1000/1270	4.18	2.83–6.18	*FAS* G-1377A AA/*FASL* T-844C CC	Chinese

[a]Meta-analyses.

versus Ala/Ala) (OR = 3.26, 95% CI 1.55–6.83), whereas males were not at increased risk. Likewise, the interaction between *SOD2* and *MPO* genotype was significant in females ($P = 0.049$) but not in males ($P = 0.75$). The gender difference was hypothesized to be related to estrogen action on *MPO* A promoter activity.

Interactions between variants in genes coding for NAT2 and mEH in determining lung cancer risk were studied by Zhou *et al.* (56). Both these enzymes have dual roles; NAT2 activates certain arylamine metabolites in cigarette smoke and deactivates aromatic amines, whereas mEH activates some PAHs and deactivates arene, alkene and aliphatic epoxides. This study found significant interactions between the *NAT2* variants associated with acetylation levels (rapid or slow) and *mEH* variants associated with activity levels (high, low, very low) and risk of lung cancer. Among non-smokers, a 50% decreased risk of lung cancer was seen for the combined *NAT2* slow-metabolizing and *mEH* high-activity genotype and a 2-fold increased risk of lung cancer was seen among subjects with 120 pack-years exposure and *NAT2* slow-metabolizing and *mEH* high-activity genotype. This complex relationship between genotypes and smoking further illustrates the complicated nature of host susceptibility to lung cancer.

DNA repair

Individual variability in DNA repair capacity may contribute to inherited susceptibility to lung cancer, with individuals who are unable to repair DNA damage or who do so at a slower rate, accumulating mutations that may modulate risk. Tobacco smoke contains many carcinogens and reactive oxygen species that produce DNA adducts, cross-links, DNA damage and DNA strand breaks requiring repair through multiple pathways, including the following: base excision repair (BER), nucleotide excision repair (NER), mismatch repair, single-strand break and double-strand break mechanisms. Phenotypic studies of DNA repair capacity have been thoroughly reviewed in Spitz *et al.* (57). In these studies, measures of total DNA repair capacity reflect variation across repair pathways; however, they are more difficult and costly to conduct than genotyping studies and usually reflect repair capacity in non-target tissue. Polymorphisms in genes coding for DNA repair enzymes active in BER (*XRCC1*, *OGG1*), NER (*ERCC1*, *XPD*, *XPA*), double-strand break repair (*XRCC3*) and mismatch repair pathways (58,59) have been studied in relation to lung cancer risk (57,60,61). For example, in the last 10 years, 27 studies of *XRCC1* including an average of 385 cases and 533 controls, 18 studies of *XPC/ERCC1* including an average of 270 cases and 380 controls and 27 studies of *XPD/ERCC2* including an average of 308 cases and 395 controls have been published (PubMed, English language, original journal articles). The largest studies are described and Table III presents results from pooled and meta-analyses. DNA repair also plays a role in response to the chemotherapeutic agent cisplatin and to ionizing radiation. Studies evaluating DNA repair polymorphisms as predictors of survival will also be discussed.

A meta-analysis performed on 13 studies of *XRCC1* found no association between the Arg194Trp, Arg280His or Arg399Gln variants and lung cancer risk (60). A large European case–control study (2188 cases and 2198 controls) also reported no association with *XRCC1* and risk of lung cancer overall (62). However, among individuals with the highest quartile of pack-years exposure, the Arg194Trp and Arg280His variants were both associated with a reduced risk of lung cancer (OR = 0.65, 95% CI 0.46–0.93, and OR = 0.56,

Table III. Risk associated with genetic polymorphisms in DNA repair enzymes

Author	No. of studies (cases/controls)	OR	95% CI	Risk genotype	Comments
XRCC1					
Hung *et al.*[a] (62)	13 (3253/3371)			Arg194Trp, Arg280His and Arg399Gln	No association of any of these polymorphisms with lung cancer risk
XPD A751C (Lys751Gln)					
Hu *et al.*[a] (65)	9 (3725/4152)	1.21	1.02–1.43	751CC versus AA	No between-study heterogeneity even with multiple ethnic groups represented
		1.19	1.02–1.40	751CC versus AC + AA	
		1.27	1.04–1.56	312AA versus GG	
		1.32	1.09–1.59	312GA + AA versus GG	
Benhamou *et al.*[a] (66)	9 (2886/3085)	1.43	1.11–1.83	312AA versus GG	US studies only
		1.25	1.03–1.52	751CC versus AA	US studies only
OGG1 Ser326Cys					
Hung *et al.*[a] (62)	8 (3252/3371)	1.24	1.01–1.53	Cys/Cys versus Ser/Ser	One study with significant heterogeneity excluded from analysis

[a]Meta-analyses.

95% CI 0.36–0.86, respectively). A tobacco–gene link was also seen in a large case–control study (1091 cases and 1240 controls) in the USA that found an interaction between tobacco use and the *XRCC1* Arg299Gln Gln/Gln genotype, with an increased risk of lung cancer in non-smokers with the variant genotype (OR = 1.3, 95% CI 1.0–1.8) and a decreased risk in heavy smokers with the variant genotype (OR = 0.5, 95% CI 0.3–1.0) (63). *XRCC1* polymorphisms may also influence response to ionizing radiation by altering the capacity to remove DNA adducts and enhance radiation sensitivity. Yoon *et al.* (64) report that progression-free survival is decreased in non-small-cell lung cancer (NSCLC) cases carrying the variant allele at Arg194Trp and receiving radiation therapy (Hazard ratio (HR) = 1.61; $P = 0.034$). Haplotypes other than TGG at the three *XRCC1* polymorphisms were associated with improved progression-free (HR = 0.32, 95% CI 0.14–0.73) and overall survival (HR = 0.20, 95% CI 0.08–0.48).

A meta-analysis examining *XPD/ERCC2* found that individuals with the 751CC genotype were at 20% increased risk of lung cancer compared with those carrying the 751AA genotype and compared with individuals carrying the AC + AA genotypes (65). Comparable 30% increased risk was also associated with the 312AA versus GG genotype and with the 312A allele (GA + AA) versus the GG genotype. Another meta-analysis of *XPD/ERCC2* polymorphisms found no association with lung cancer risk and the Asp312Asn (G312A) or the Lys751Gln (A751C) polymorphisms across all studies (66). Significant findings were seen when the analysis was limited to those studies conducted in the USA (Asn/Asn versus Asp/Asp: OR = 1.43, 95% CI 1.11–1.83; Gln/Gln versus Lys/Lys: OR = 1.25, 95% CI 1.03–1.52). This variation in findings by geographic region may be due to differences in variant allele frequencies across populations. *XPD* variants may also alter removal of platinum-DNA adducts, thereby altering sensitivity to platinum-based chemotherapeutic agents. In a small study of 108 NSCLC patients, Booten *et al.* (67) report poorer survival in patients carrying the asparagine 312/glutamine 751 haplotype compared with patients carrying the aspartic acid 312/lysine 751 haplotype. In another study of 109 patients, Ryu *et al.* (68) showed no association between *XPD* polymorphisms and survival.

OGG1 Ser326Cys was the focus of a meta-analysis including eight studies. A 24% increased risk of lung cancer in individuals carrying the *OGG1* Cys/Cys genotype compared with the Ser/Ser genotype was reported (60). A large, multi-center European study including 2188 cases and 2198 controls also reported an increased risk of adenocarcinoma of the lung associated with the *OGG1* Cys/Cys versus Ser/Ser genotype (OR = 1.66, 95% CI 1.04–2.66) (62). This association was not detected for squamous cell carcinomas and did not vary by smoking status or age at diagnosis, although limited sample size in some of these strata limit the power to detect such an association.

In a smaller study (343 cases and 413 controls), Zienolddiny *et al.* (69) examined a panel of 44 single-nucleotide polymorphisms (SNPs) in 20 DNA repair genes. Polymorphisms in genes in *ERCC1*, *ERCC2*, *ERCC5*, *XPA*, *APE1/APEX*, *OGG1*, *PCNA*, *XRCC1*, *ATR*, *NSB1*, *XRCC2* and *XRCC9* were all associated with risk of NSCLC in this group of smokers. Variation in risk occurred with pack-years of exposure. Functional status was not always known for the SNPs included and no analysis of haplotypes or combinations of SNPs was included in this study.

The NER enzyme ERCC1 removes bulky intra-strand platinum-DNA adducts formed by platinum-based chemotherapeutic agents. In a study of 761 patients set within a clinic trial, ERCC1 protein expression in lung tumors was evaluated (70). Among participants receiving cisplatin-based adjuvant therapy, survival was improved in those with ERCC1-negative tumors (HR = 0.65, 95% CI 0.50–0.86). Ryu *et al.* (68) report that the non-variant CC genotype at the synonymous SNP Asn118Asn is associated with improved survival, but this study only included 109 patients.

Gene–gene combinations/interactions. Like carcinogen-metabolizing candidate gene studies, the combined effects of multiple DNA repair pathway gene polymorphisms have been evaluated (Table II).

Both XRCC1 and adenosine diphosphate ribosyl transferase (ADPRT) function in the BER pathway. Mechanistically, ADPRT binds to DNA strand breaks, recruiting other enzymes such as XRCC1 to initiate the BER process. One large hospital-based study in China (71) found a significant interaction between the *ADPRT* Val762Ala and *XRCC1* Arg399Gln Gln/Gln genotypes ($P = 0.018$), resulting in a close to 6-fold increased risk in those carrying the *ADPRT* Ala/Ala low-activity risk allele and *XRCC1* Gln/Gln low-activity risk allele. The subset of individuals carrying this high-risk genotype, however, was very small (2.0% of the cases and 0.5% of the controls). These findings in the Chinese population also may not apply to the US population, as allele frequencies differ between the two populations; the *ADPRT* 762Ala allele frequency is 13.7% for the Han Chinese, 2% among US Caucasians and 0% among US African Americans; the *XRCC1* Gln/Gln genotype frequency is 8.9% for the Han Chinese and 11.5% in a Caucasian population (63). In addition, the Zhang *et al.* study did not adjust for smoking amounts or pack-year levels when examining the gene–gene interactions to quantify the burden of tobacco smoke on low-activity DNA repair variants.

Zhou *et al.* (63) evaluated genes in the BER and NER DNA repair pathways in Caucasians to examine the interactions between *XRCC1* Arg399Gln, *ERCC2 (XPD)* Asp312Asn and Lys751Gln genotypes and cigarette smoking. The *XRCC1* Gln/Gln genotype was associated with a 2.4-fold increased risk of lung cancer among non-smokers, but showed a protective effect on lung cancer risk among heavy smokers

(OR = 0.5, 95% CI 0.3–1.0). The combined effect of polymorphisms in *XRCC1* and *ERCC2* was evaluated by grouping individuals into six categories based on increasing number of risk alleles. Risk of lung cancer increased with increasing number of risk genotypes among non-smokers. Risk associated with number of risk genotypes declined with increasing amount smoked. The percentage of individuals carrying six risk alleles was very small (≤1.5%, except in non-smoking cases where it was 4.1%).

Inflammation-related genes

Inflammation has been thought to play a role in carcinogenesis of a variety of cancers, including lung cancer (72). Chronic inflammation may be due to infectious agents, such as *Helicobacter pylori*, hepatitis C and the Epstein–Barr virus, or the result of other environmental exposures such as asbestos, silica and tobacco smoke (72). Epidemiological evidence suggests that a prior diagnosis of lung disease associated with inflammation, including COPD, pneumonia and tuberculosis, is a risk factor for later lung cancer development (73,74) and data from animal models support a link between inflammation and tumor incidence and growth (75,76). The potential mechanisms linking chronic inflammation and cancer are diverse and well-described by Coussens *et al.* (72).

Genes involved in the regulation of the inflammatory response are recent additions to the lung cancer candidate gene literature. The small body of epidemiologic work in this area has focused on polymorphisms in genes that encode interleukins (ILs), matrix metalloproteinases (MMPs) and cyclooxygenase-2 (COX2).

ILs are cytokines that are involved in both innate and adaptive immunity. Polymorphisms in the genes that code a particular IL or IL receptor may influence immune response and ultimately individual susceptibility to cancer. Two polymorphisms in the promoter region of the *IL1β* gene (T-31C and C-511T) have been associated with increased risk of lung cancer in 251 NSCLC patients and 272 controls from Norway (-31TT: OR = 2.4, 95% CI 1.3–4.4; -511CC: OR = 2.5, 95% CI 1.5–4.6) (77). Further investigation suggests that a tandem repeat polymorphism in the interleukin-1 receptor antagonist (*IL-1Ra*) gene in combination with the *IL1β* C-31T polymorphism may also increase susceptibility to lung cancer (78). *IL8* T-251A and *IL6* G-174C promoter polymorphisms have also been associated with lung cancer risk in females (-251AA genotype: OR = 0.2, 95% CI 0.06–1.0) and squamous cell carcinomas (-174CC genotype: OR = 2.2, 95% CI 1.10–4.68), respectively (79). These subsets came from a case–control study including 250 cases and 214 controls and each stratum includes <50 individuals. These results have yet to be replicated in other populations and dozens of ILs and other molecules associated with the inflammatory response remain to be explored.

MMPs are a family of zinc metalloproteases that are responsible for degradation of the extracellular matrix and thus are thought to play a role in angiogenesis, cell proliferation and apoptosis (80) involved in repair during inflammation. Although therapies targeting MMPs have failed to meet expectations (81), some research suggests that certain MMPs may be associated with increased risk of lung cancer. Research by Su *et al.* (82,83) suggests that polymorphisms in *MMP-1* (-1607 1G/2G) and certain haplotypes of functional polymorphisms in *MMP-1*, *MMP-3* and *MMP-12* lead to increased risk of lung cancer in non-smokers and men. *MMP-1* 2G/2G genotype was not associated with lung cancer risk in all individuals (1752 cases and 1363 controls) as compared with the 1G/1G genotype (OR = 1.14, 95% CI 0.90–1.45). When additional *MMPs* were evaluated and haplotypes included, no individual *MMP* was associated with lung cancer risk and the number of *MMP* variant alleles was not associated with lung cancer risk. An *MMP-2* promoter SNP (C-1306T) has been associated with a 2-fold increase in lung cancer risk (95% CI 1.7–2.8) in a study of 781 cases and 852 controls (84). The *MMP-2* genotype–lung cancer risk association was stronger as cigarette smoke exposure increased. These findings need to be confirmed in other populations but suggest that *MMPs* are another potentially important candidate pathway in lung carcinogenesis, with variation in risk by amount smoked. MMPs may also serve as prognostic markers because of their role in degradation of the extracellular matrix leading to tumor invasion and metastasis. Heist *et al.* (85) evaluated SNPs in *MMP-12*, *MMP-1*, *MMP-2*, and *MMP-3* in 382 NSCLC patients with stage 1 disease who were resected. Overall survival was associated with the *MMP-12* 1082A/G polymorphism. Individuals carrying the variant allele had poorer overall survival (HR = 1.94, 95% CI 1.28–2.97).

A third group of enzymes, COXs, are also involved in the inflammatory process. Two COX isoforms have been identified, COX1 and COX2. *COX2* is over-expressed in many tumor types, including lung cancers, and has been identified as a marker of poor prognosis in non-small-cell lung cancers (86–88). Additionally, use of non-steroidal anti-inflammatory drugs, which target COX2, has been shown to be chemopreventive in lung cancer, particularly among smokers (89). A small study of 250 lung cancer patients and 214 controls in Norway described 2-fold and 4-fold increased risk among heterozygotes and homozygotes carrying the C/T (95% CI 1.25–3.59) and T/T (95% CI 2.44–7.49) genotype in the 3′-untranslated region of the COX2 gene (79,90). The biological role of this polymorphism remains unclear, however, inhibiting COX2 via non-steroidal anti-inflammatory drug therapy is undergoing clinical trials to determine if altering this pathway may increase survival from lung cancer or decrease its occurrence (91). Understanding modulation of the inflammation pathway will require larger studies focused on multiple genes in this process.

Cell cycle genes

Numerous enzymes play a role in keeping the cell cycle in check, including p53 (TP53), p73, p15, p16, p21, mouse double minute 2 (MDM2), FAS and FASL. A large Chinese population was the focus of one study exploring the relationship between two cell cycle genes, *MDM2* and *TP53*, and risk of lung cancer (Table II) (92). The polymorphisms *MDM2* G-5 + 309T and *TP53* Arg72Pro were, individually, associated with lung cancer risk. Additionally, individuals carrying the *MDM2* GG and *TP53* Pro/Pro genotypes were at 4.6-fold increased risk of lung cancer risk (95% CI 2.76–7.54; *P* for interaction <0.05). A significant gene–gene-smoking interaction was also present; the highest risk occurred among smokers with both *MDM2* GG and *TP53* Pro/Pro genotypes (OR = 10.41, 95% CI 5.26–20.58).

Another study explored the combined effects of *p53* (G72C, intron 3 duplication and intron 6 G to A transversion) and *p73* (G4C14 to A4T14) polymorphisms on lung cancer risk (93). The combined *p53* and *p73* variant alleles were associated with a modest increase in risk of lung cancer (OR = 1.13, 95% CI 1.05–1.21), using the combined number of variant alleles as a continuous variable. With one to three variants, lung cancer risk increased 30% and with greater than or equal to four variants the risk increased 78%. Similar trends were seen among former smokers, among subjects over age 66 and among men, whereas only greater than or equal to four variant alleles increased risk of lung cancer for current smokers and subjects younger than age 59.

A large study in the Han Chinese population examined gene–gene interactions in a pro-apoptotic pathway: the FAS (receptor) and FASL (ligand) system (94). Zhang *et al.* found that individually, subjects with the *FAS* G-1377A AA or the *FASL* T-844C CC genotypes were at increased risk of lung cancer (OR = 1.59, 95% CI 1.21–2.10, and OR = 1.79, 95% CI 1.26–2.52, respectively), whereas those with the combined *FAS* AA and *FASL* CC genotype had a 4.18-fold increased risk (95% CI 2.83–6.18). Their results suggest a multiplicative interaction between these genes.

The prognostic significance of somatic *TP53* mutations and expression has been studied by many (95–98). Nelson *et al.* (99) have shown that the combination of a somatic mutation in *PT53* and carrying a codon 72 proline allele is associated with worse survival outcomes (HR = 2.6; $P < 0.03$).

Genes and intermediate endpoints

DNA adducts. As an intermediate step in the study of genetic polymorphisms and risk of lung cancer, the relationship between SNPs

and DNA adducts has been studied, focusing on polymorphisms within genes in the metabolic activation/detoxification (*GSTM1*, *GSTT1*, *NAT2*, *CYP1A1*, *CYP2D6*, *CYP2E1*), tumor suppression (*TP53*), inflammation (*ILRn*, *IL1B*), and DNA repair (*XPD*, *XRCC1*, *XRCC3*, *hOGG1*, *hNTH1*) pathways. The effects of these genetic variants have been measured using various indicators of adduct formation, including bulky adduct levels in peripheral lymphocytes, normal lung tissue, normal nasal mucosa and normal and tumor bronchial tissue, and also by the level of somatic mutations in the *HPRT* and *TP53* genes.

Within the metabolic pathway genes studied to date, an independent association between *GSTM1* genotype and adduct levels has been observed by several studies (19,100–102). Shields *et al.* (19) reported significant associations between the *GSTM1*-null genotype and both the presence of detectable PAH-DNA adducts ($P = 0.04$) and the level of PAH-DNA adducts (0.20 fmol/μg in *GSTM1* null versus 0.038 fmol/μg in *GSTM1* present, $P = 0.02$) in 38 lung tissue samples. Similarly, Kato *et al.* (99) observed an association between *GSTM1*-null genotype and the presence of detectable PAH-DNA adducts in lung tissue samples from 90 cancer-free autopsy donors; however, this group did not report PAH-DNA levels by *GSTM1* genotype. Results from Ryberg *et al.* (100) also suggest higher mean DNA adduct levels in individuals with a *GSTM1*-null genotype than in those with a *GSTM1*-present genotype (1.5 ± 8.5 versus 9.4 ± 6.0 adducts per 10^8 nucleotides, $P = 0.088$) in 70 lung cancer patients. In contrast, other studies have not observed an association between *GSTM1* genotype and DNA adducts (102–108).

Two studies report an independent association between *CYP1A1* variant alleles and adduct levels (102,105); however, most studies have not provided evidence for an association of *CYP1A1* polymorphisms with adducts (19,100,106,107,109). Of these negative studies, all had <100 subjects included with the exception of the report by Schoket *et al.* (108), which found no association between *CYP1A1* genotype and adduct level in bronchial tissue in 150 pulmonary surgery patients (126 with lung cancer). A study that included *CYP2D6* and *CYP2E1* genotypes reported significant differences in lung tissue mean adduct levels for both loci ($P = 0.01$ and $P = 0.05$, respectively) in 90 autopsy donors (100). The *GSTP1* AA genotype was associated with lower DNA adduct level than the AG genotype ($P = 0.01$) and the GG genotype ($P = 0.02$) in 70 lung cancer samples in a study examining this gene (100).

Several gene combinations in the metabolic pathway have been examined, and significant associations with adduct levels have been observed for several of these. In two different Swedish populations, Hou *et al.* (103,104) reported higher peripheral lymphocyte DNA adducts in the *GSTM1*-null/*NAT2*-slow group as compared with individuals with the *GSTM1*-present/*NAT2*-fast genotype ($P = 0.03$). The second study, which included 185 lung cancer cases and 164 controls and examined this combination by both case–control status and smoking status, observed an increase in DNA adduct level (from peripheral lymphocytes) among currently smoking controls with the *GSTM1*-null/*NAT2*-slow genotype compared with individuals with *GSTM1*-present/*NAT2*-rapid genotypes (OR = 19.3, 95% CI 1.1–338.6), and an increase in *HPRT* mutations in ever-smoking cases and controls with this genotype and lower pack-years of smoking (OR = 3.7, 95% CI 1.3–10.7). The combination of *GSTM1*-null and *CYP1A1*-homozygous mutant was associated with higher adduct levels ($P = 0.017$) in DNA derived from lung tissue from 20 lung cancer patients and peripheral lymphocytes from 20 coke oven workers (102). *GSTM1* null in combination with *GSTP1* AG or GG genotypes has also been associated with higher lung tissue DNA adduct levels, as compared with all other *GSTM1*/*GSTP1* genotype combinations ($P = 0.011$) by Ryberg *et al.* (101) in 70 current-smoking lung cancer patients.

The most commonly studied DNA repair gene in relation to DNA adducts is *XPD*. Three studies observed a positive association between adduct levels in lymphocytes and the *XPD* 751Gln polymorphism in exon 23 (110–112). In a sample of 171 Swedish lung cancer cases and 146 controls, Hou *et al.* (110) observed an association between peripheral lymphocyte adduct level and the number of *XPD* 751Gln alleles ($P = 0.005$) and statistically significant associations between the number of 312Asn variant alleles alone ($P = 0.01$) and in combination with the 751Gln polymorphism ($P = 0.02$). A study by Mechanic *et al.* (113) examined the relationship between both these *XPD* polymorphisms and somatic *TP53* mutations in 309 lung tumors, but did not observe any associations, either singly or in combination. Palli *et al.* (111) reported an association between the *XPD* 751Gln allele and peripheral lymphocyte DNA adduct levels ($P = 0.02$); no association between two additional DNA repair genes, *XRCC1* Arg399Gln or *XRCC3* Thr241Met, and adduct levels was observed. Matullo *et al.* (112) compared high (>median) versus low peripheral lymphocyte DNA adduct level in 308 healthy participants and found that the homozygous *XPD* 751Gln genotype was associated with high adduct level, but only in the never-smokers (OR = 3.81, 95% CI 1.02–14.16). In this same sample, associations between having the *XRCC1* 399Gln allele (never-smokers only) or *XRCC3* 241Met allele and higher adduct level were also observed ($P = 0.007$ and $P = 0.04$, respectively). Zienolddiny *et al.* (69) evaluated a panel of 44 SNPs in 20 different DNA repair genes in relation to PAH-DNA adducts in normal lung tissue of 211 lung cancer patients. Greater than the mean number of adducts was seen in individuals carrying the *XRCC1* Arg194Trp variant ($P = 0.015$). The *XRCC1* Arg280His, *XRCC1* Arg399Gln, *ERCC1* G8092T, *ERCC5* His46His and *MGMT/AGT* Lys178Arg variant genotypes were associated with lower than the mean number of adducts.

One study of inflammatory pathway genes in 209 Norwegian individuals with lung cancer suggested a possible role of *IL1RN* in adduct level, as *ILRN**1 carriers had a nearly 2-fold higher level of DNA adducts in normal lung tissue than non-carriers ($P = 0.057$); no association was observed for *IL1B1* mutations (78). The reported associations between genetic polymorphisms and adduct levels in lung cancer, however, should be interpreted with caution, as there is little or no replication of results to date, and most studies have been small.

Epigenetic events

Although the role of inherited variation in DNA sequence is not yet fully understood, it is likely to modify susceptibility to lung cancer. Acquired factors not directly encoded within the DNA sequence may also play a role. Epigenetic events, such as DNA methylation, histone deacetylation and phosphorylation, influence DNA expression and thus may affect carcinogenic potential. Several studies have focused on associations between epigenetic events and lung cancer risk. Global hypomethylation may result in genomic instability primarily through loss of methylation in repetitive genomic regions. Methylation changes may also occur as a result of disregulation of DNA methyltransferase I, an enzyme responsible for maintaining methylation patterns. The most widely studied epigenetic event in relation to lung cancer is regional hypermethylation. Methylation of CpG islands in promoter regions acts to silence genes and affects chromosomal stability. Genes in multiple pathways have been investigated and include cell cycle genes (*p16*), apoptosis genes (*DAPK* and *RASSF1A*), cell differentiation and proliferation genes (*RAR-β*) and DNA repair genes (*MGMT*). Paired normal and tumor tissues, tissues from cases and controls and surrogate biospecimens have been used to study methylation patterns.

p16 is methylated in ~25–41% of NSCLC, *RAR-β* in 40–43%, *RASSF1A* in 30–40%, *DAPK* in 16–44% and *MGMT* in 16–27% (114,115). Methylation of *p16* is seen as an early event in the carcinogenic process and can be identified in sputum before diagnosis of lung cancer (116). Variation in methylation status has been associated with cigarette smoke exposure (117–119). Several excellent reviews have been published (120–122); therefore, this review will only touch on studies that report on variation in methylation patterns associated with candidate susceptibility genes. In addition, varying associations of *p16* methylation with prognosis have been observed. *p16* methylation has been associated with poorer prognosis after a diagnosis of NSCLC (123), associated with poorer prognosis in combination with

hypermethylated *FHIT* (124) and not associated with prognosis (125,126).

A cross-sectional study of 70 former underground uranium miners at high risk of lung cancer, but cancer free at the time, examined the relationship between polymorphisms in *XRCC1*, *GSTM1*, *GSTP1*, *NQO1*, and *MPO* and aberrant methylation of *p16* and *MGMT* in the sputum (126). *MGMT* methylation was associated with the *GSTP1* Ile/Val or Val/Val genotype (OR = 4.8, 95% CI 1.2–18.6) and increased *MGMT* or *p16* methylation was associated with the *NQO1* CT/CC genotype (OR = 3.1, 95% CI 1.0–9.2) and the *GSTP1* Ile/Val or Val/Val genotype (OR = 4.4, 95% CI 1.3–14.2), all adjusted for ethnicity. Chan *et al.* (128) evaluated the relationship between lung cancer susceptibility polymorphisms and aberrant methylation in a series of 75 lung cancer patients. *p16* methylation was more likely in individuals carrying the *MPO* G/A or A/A genotype than in those carrying the G/G genotype (OR = 3.2, 95% CI 1.20–8.53). The same was true for those with the *XRCC1* CT or TT genotype (OR = 4.47, 95% CI 1.34–14.93). The probability of *RAR-β* methylation was increased for those with the *XRCC1* CT or TT genotype (OR = 7.67, 95% CI 1.62–36.18). Methylation patterns at multiple loci, in normal tissues, lung tumors and surrogate biospecimens, should be characterized to more fully appreciate the epigenetic progression leading to development of cancer. Although only relatively small studies have been conducted to date, methylation may serve as an important early diagnostic and prognostic marker and a reversible target for treatment.

Conclusions and future directions

The aim of profiling the genetic and epigenetic characteristics of an individual at high risk for lung cancer is ongoing. In spite of hundreds of candidate gene studies, the specific genes that are responsible for enhanced risk remain poorly understood. Risks associated with some of the mechanistically plausible candidates described are still uncertain after a number of population studies and others remain to be detected. Single-gene studies conducted to date have a number of limitations, which have contributed to inconclusive results including small sample size and associated low power to detect moderate risks when allele frequencies are low. False-positive results are a potential problem and more likely when initially small data sets are stratified by age, race, gender and smoking history. Movement towards the development of consortia to pool findings across studies and increase sample size and power will address some of these issues. Several consortia are in place including GSEC and the International Lung Cancer Consortium. Although these groups will be able to assemble large numbers of cases and controls already genotyped at a variety of SNPs, there will be some limitations including population heterogeneity due to significant differences in allelic frequencies between races, differing case and exposure definitions and differing genotyping methods. Overall, the GSEC collaboration has organized almost 230 genetic studies to acquire thousands of subjects for gene–gene interaction analyses (129) and International Lung Cancer Consortium expects to have 25 000 cases and 25 000 controls represented, allowing for studies of gene–environment and gene–gene interactions, as well as addressing risk in never-smokers and in those with early onset disease, and by histologic type. Pooled analyses are more flexible, allowing for the creation of standard variable and risk genotype definitions, adjustment for confounders, analyses by subgroup and evaluation of interactions (130). There are also limitations to pooled analyses. They are subject to population heterogeneity resulting from variation in allele frequencies across study populations. Genotyping a series of ancestry informative markers and adjustment for individual ancestry can be used to reduce this potential bias, but most pooled analyses will not have these data available. Alternatively, population heterogeneity may be reduced by limiting pooled analyses to populations with the same allele frequencies for the genetic variant under study. The meta-analyses presented also have limitations, most notably the potential for publication bias due to the low probability that studies with null results are published resulting in an overestimation of risk (130). Meta-analyses also are limited to use of estimates that are obtained after adjustment for different sets of confounders and use of different statistical methods. Analyses of gene–environment or gene–gene interactions or stratified analyses are not possible using this approach. Both pooled and meta-analyses should include a rigorous evaluation of heterogeneity between studies and report with caution when such heterogeneity exists (131).

With funding support and a willingness of investigators to share stored DNA, collaborations could move the field forward by carefully selecting candidate genes and SNPs within those genes for further study. For the majority of sporadic cancers, the proposed genetic susceptibility model is one where multiple genes are acting, each conferring moderate increases in risk. The SNPs that have been studied so far represent only some of the variation within a gene, may not be functional and are unlikely to be acting alone. Selection of candidate genes within pathways and genotyping at multiple markers within a gene are approaches that are gaining support. Marker selection within a gene can be directed at a set of tagSNPs, based on haplotype structure representing most of the variation within a gene, and/or can include only those SNPs with known functionality. These efforts have been greatly facilitated by the HapMap project. The goals of this project are to optimize SNP selection to identify the minimum number of SNPs (tagSNPs or tSNPs) representing genetic variation across a chromosomal region for use in association studies (132). There are some limitations to use of tagSNPs identified through the HapMap project (133). TagSNPs were designed with minor allele frequencies of >5% and were identified in a relatively small number of samples from four populations (Utah representing northern and western Europeans, Han Chinese, Japanese and Nigerians) effectively excluding rare variants and potentially discounting allelic heterogeneity. The capture of common genetic variation using the HapMap has been shown in non-African American samples (134), in Estonians (133), in the Spanish population (135) and in the Finnish population (136). There is also the question of whether the HapMap SNPs can capture variation across untyped SNPs. Tantoso *et al.* (137) compared variation captured by HapMap SNPs with that captured by National Institute of Environmental Health Sciences SNPs, which included SNPs identified by re-sequencing. The results show that HapMap SNPs capture ~55% of the untyped variants in European and Asian populations, but only 30% in African populations. Variability in coverage was related to SNP density. Montpetit *et al.* (133) suggest that a SNP density of <5 kbp, as available in Phase I of the HapMap, may not be optimal for selection of tagSNPs. The Phase II map will increase SNP density and improve coverage of tagSNPs. Re-sequencing of candidate genes to identify rare SNPs and obtain more complete coverage in candidate regions, therefore, may also be needed. Even with the caveats noted in SNP selection, the use of new technology to genotype thousands of candidate SNPs per sample will allow for more complete coverage of the variation within candidate genes in multiple pathways.

Whole-genome association studies have been suggested as an alternative to the candidate gene approach for the discovery of genes for complex diseases. Hundreds of thousands of closely spaced SNPs are used in a genome-wide association study, with priority given to non-synonymous SNPs. For modest relative risks and where environmental factors play a role, association studies are more powerful across a range of genotype frequencies than linkage studies (138). Genotyping hundreds of thousands of SNPs spanning the genome is now feasible, but still costly. The whole-genome approach can identify multiple genes or regions that can then be followed up with specific candidate gene analyses. The approach will be relevant even if a specific candidate were already known, as known candidate genes will probably account for a modest proportion of excess risk overall. The advantages of this approach include no assumptions as to gene function or mode of inheritance, families are not needed, common genetic variants with modest effects can be detected, no variation in allele frequencies between populations is needed (as is true for admixture mapping) and the analysis is straightforward. Currently, choice of SNPs for whole-genome association studies is limited by cost constraints and available commercial panels. These panels typically include tagSNPs selected using the HapMap and non-synonymous

SNPs that result in amino acid changes. This approach requires extremely large sample sizes and can be made more efficient by focusing on cases most likely to have a genetic component to risk, i.e. those with a family history of lung cancer or non-smokers (139). To date, no such comprehensive, whole-genome association studies for lung cancer have been completed.

Admixture mapping provides an alternative to whole-genome association studies when risk allele frequencies and disease risk vary by race/ethnicity (140). This approach depends on the existence of measurable linkage disequilibrium between markers and disease loci and the extent of linkage disequilibrium present throughout the genome of a population. Recently admixed populations, such as African Americans, provide the most power for this type of analysis. Although not detailed in this review because studies include small sample sizes, risks in African Americans associated with common candidate susceptibility genes often are greater than those in Caucasians and may indicate increased susceptibility in this population (141–143). In addition, familial aggregation of lung cancer is stronger in African Americans than in Caucasians (144). Focusing studies of genetic risk in race/ethnic-specific populations should be considered.

The goal of all these approaches is to identify a set of genetic and/or molecular predictors of lung cancer risk and prognosis. Translation of results from these molecular epidemiology studies into use in the clinical setting for screening, early detection, treatment development and prevention will require proof of sensitivity, specificity and positive predictive value. The National Cancer Institute's Early Detection Research Network has developed criteria for assessing new molecular and genetic markers ready for validation studies (145). These criteria include the evaluation of biologic rationale and strength of hypotheses; strength of design; technical parameters such as reproducibility, sensitivity and specificity of the assay/genotype; clinical or scientific impact and practicality. Validation of epidemiologic findings typically relies on consistency across study populations. Most studies of the genetic/molecular contributions to risk and prognosis, at the current state of the science, are directed at understanding mechanisms of lung carcinogenesis and are unlikely, in isolation, to result in reductions of lung cancer occurrence in the short term. However, as lung cancer risk or prognosis profiles are developed, issues related to sensitivity, specificity and application of the marker panel as a screening tool in the clinical setting will need to be addressed (146).

There are strengths and limitations to the studies and approaches described and the 'best' approach to the identification of genetic and molecular changes associated with lung cancer risk and prognosis is unclear. The age of single SNP studies of candidate genes is past. Where candidate genes have been hypothesized, the use of tagSNPs, selected both through existing databases and re-sequencing, offers the most cost-effective approach. Epigenetic alterations within candidate genes need to be considered as an alternative mechanism of variation. Given the complexity of the genome and pathways involved in carcinogenesis, it is probable that there are pathways yet to be fully characterized. New leads will come through genome-wide association studies, at great cost, with concern for confidentiality and need for cooperation across institutions.

Identification of a subset of the population, in particular smokers, who are at highest risk of developing lung cancer, is important for many reasons. A high-risk population needs to be defined and used as the target for screening studies. Without a set of genes identified to define the highest risk group, family history can be used as a risk factor, in addition to smoking history and pulmonary function, for entry into screening trials. A characterized high-risk population, both in terms of genotype and epigenetic measures, also should be the target for chemoprevention and treatment trials. The specific genes associated with high risk will need to be identified more thoroughly to reach this goal.

Acknowledgements

This work was supported in part by National Institutes of Health grants R01 CA60691, R01 CA87895 and P30 CA22453 and contract N01 PC35145.

References

1. Ries,L.A.G. *et al*. and (eds) (2004) SEER Cancer Statistics Review, 1975-2001. http://seer.cancer.gov/csr/1975_2001/. National Cancer Institute, Bethesda, MD.
2. Jemal,A. *et al*. (2006) Cancer statistics, 2006. *CA Cancer J. Clin.*, **56**, 106–130.
3. Doll,R. *et al*. (1952) A study of the aetiology of carcinoma of the lung. *Br. Med. J.*, **2**, 1271–1286.
4. Doll,R. *et al*. (1976) Mortality in relation to smoking: 20 years' observations on male British doctors. *Br. Med. J.*, **2**, 1525–1536.
5. Hammond,E.C. *et al*. (1980) Smoking and cancer in the United States. *Prev. Med.*, **9**, 169–174.
6. Mattson,M.E. *et al*. (1987) What are the odds that smoking will kill you? *Am. J. Public Health*, **77**, 425–431.
7. Bromen,K. *et al*. (2000) Aggregation of lung cancer in families: results from a population-based case-control study in Germany. *Am. J. Epidemiol.*, **152**, 497–505.
8. Matakidou,A. *et al*. (2005) Systematic review of the relationship between family history and lung cancer risk. *Br. J. Cancer*, **93**, 825–833.
9. Mayne,S.T. *et al*. (1999) Familial cancer history and lung cancer risk in United States nonsmoking men and women. *Cancer Epidemiol. Biomarkers Prev.*, **8**, 1065–1069.
10. Ooi,W.L. *et al*. (1986) Increased familial risk for lung cancer. *J. Natl Cancer Inst.*, **76**, 217–222.
11. Schwartz,A.G. *et al*. (1996) Familial risk of lung cancer among nonsmokers and their relatives. *Am. J. Epidemiol.*, **144**, 554–562.
12. Tokuhata,G.K. *et al*. (1963) Familial aggregation of lung cancer in humans. *J. Natl Cancer Inst.*, **30**, 289–312.
13. Bailey-Wilson,J.E. *et al*. (2004) A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am. J. Hum. Genet.*, **75**, 460–474.
14. Kawajiri,K. *et al*. (1990) Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. *FEBS Lett.*, **263**, 131–133.
15. Bartsch,H. *et al*. (2000) Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 3–28.
16. Hirvonen,A. *et al*. (1992) Metabolic cytochrome P450 genotypes and assessment of individual susceptibility to lung cancer. *Pharmacogenetics*, **2**, 259–263.
17. Le Marchand,L. *et al*. (1998) Associations of CYP1A1, GSTM1, and CYP2E1 polymorphisms with lung cancer suggest cell type specificities to tobacco carcinogens. *Cancer Res.*, **58**, 4858–4863.
18. London,S.J. *et al*. (2000) CYP1A1 I462V genetic polymorphism and lung cancer risk in a cohort of men in Shanghai, China. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 987–991.
19. Shields,P.G. *et al*. (1993) Lung cancer, race, and a CYP1A1 genetic polymorphism. *Cancer Epidemiol. Biomarkers Prev.*, **2**, 481–485.
20. Tefre,T. *et al*. (1991) Human CYP1A1 (cytochrome P(1)450) gene: lack of association between the Msp I restriction fragment length polymorphism and incidence of lung cancer in a Norwegian population. *Pharmacogenetics*, **1**, 20–25.
21. Houlston,R.S. (2000) CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. *Pharmacogenetics*, **10**, 105–114.
22. Le Marchand,L. *et al*. (2003) Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States). *Cancer Causes Control*, **14**, 339–346.
23. Vineis,P. *et al*. (2003) CYP1A1 T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. *Int. J. Cancer*, **104**, 650–657.
24. Taioli,E. *et al*. (2003) Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years. *Int. J. Epidemiol.*, **32**, 60–63.
25. Hung,R.J. *et al*. (2003) CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis*, **24**, 875–882.
26. Goto,I. *et al*. (1996) Prognostic significance of germ line polymorphisms of the CYP1A1 and glutathione S-transferase genes in patients with non-small cell lung cancer. *Cancer Res.*, **56**, 3725–3730.
27. Houlston,R.S. (1999) Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. *Cancer Epidemiol. Biomarkers Prev.*, **8**, 675–682.
28. Benhamou,S. *et al*. (2002) Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. *Carcinogenesis*, **23**, 1343–1350.

29. Ye,Z. *et al*. (2006) Five glutathione s-transferase gene variants in 23, 452 cases of lung cancer and 30, 397 controls: meta-analysis of 130 studies. *PLoS Med.*, **3**, e91.
30. Sweeney,C. *et al*. (2003) Glutathione S-transferase M1, T1, and P1 polymorphisms and survival among lung cancer patients. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 527–533.
31. Lu,C. *et al*. (2006) Association between glutathione S-transferase pi polymorphisms and survival in patients with advanced nonsmall cell lung carcinoma. *Cancer*, **106**, 441–447.
32. Feyler,A. *et al*. (2002) Point: myeloperoxidase –463G –> a polymorphism and lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1550–1554.
33. Le Marchand,L. *et al*. (2000) Association of the myeloperoxidase –463G–> a polymorphism with lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **9**, 181–184.
34. Schabath,M.B. *et al*. (2002) A myeloperoxidase polymorphism associated with reduced risk of lung cancer. *Lung Cancer*, **37**, 35–40.
35. Xu,L.L. *et al*. (2002) Counterpoint: the myeloperoxidase –463G–> a polymorphism does not decrease lung cancer susceptibility in Caucasians. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1555–1559.
36. Kiyohara,C. *et al*. (2005) NQO1, MPO, and the risk of lung cancer: a HuGE review. *Genet. Med.*, **7**, 463–478.
37. Chen,H. *et al*. (1999) Association of the NAD(P)H:quinone oxidoreductase 609C–>T polymorphism with a decreased lung cancer risk. *Cancer Res.*, **59**, 3045–3048.
38. Sunaga,N. *et al*. (2002) Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 730–738.
39. Wiencke,J.K. *et al*. (1997) Lung cancer in Mexican-Americans and African-Americans is associated with the wild-type genotype of the NAD(P)H: quinone oxidoreductase polymorphism. *Cancer Epidemiol. Biomarkers Prev.*, **6**, 87–92.
40. Xu,L.L. *et al*. (2001) The NAD(P)H:quinone oxidoreductase 1 gene polymorphism and lung cancer: differential susceptibility based on smoking behavior. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 303–309.
41. Benhamou,S. *et al*. (1998) Association between lung cancer and microsomal epoxide hydrolase genotypes. *Cancer Res.*, **58**, 5291–5293.
42. Cajas-Salazar,N. *et al*. (2003) Effect of epoxide hydrolase polymorphisms on chromosome aberrations and risk for lung cancer. *Cancer Genet. Cytogenet.*, **145**, 97–102.
43. Gsur,A. *et al*. (2003) Association of microsomal epoxide hydrolase polymorphisms and lung cancer risk. *Br. J. Cancer*, **89**, 702–706.
44. Lee,W.J. *et al*. (2002) Microsomal epoxide hydrolase polymorphisms and lung cancer risk: a quantitative review. *Biomarkers*, **7**, 230–241.
45. London,S.J. *et al*. (2000) Lung cancer risk in relation to genetic polymorphisms of microsomal epoxide hydrolase among African-Americans and Caucasians in Los Angeles County. *Lung Cancer*, **28**, 147–155.
46. Park,J.Y. *et al*. (2005) Genetic analysis of microsomal epoxide hydrolase gene and its association with lung cancer risk. *Eur. J. Cancer Prev.*, **14**, 223–230.
47. Wu,X. *et al*. (2001) The association of microsomal epoxide hydrolase polymorphisms and lung cancer risk in African-Americans and Mexican-Americans. *Carcinogenesis*, **22**, 923–928.
48. Zhao,H. *et al*. (2002) Microsomal epoxide hydrolase polymorphisms and lung cancer risk in non-Hispanic whites. *Mol. Carcinog.*, **33**, 99–104.
49. Zhou,W. *et al*. (2001) The interaction between microsomal epoxide hydrolase polymorphisms and cumulative cigarette smoking in different histological subtypes of lung cancer. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 461–466.
50. Kiyohara,C. *et al*. (2006) EPHX1 polymorphisms and the risk of lung cancer: a HuGE review. *Epidemiology*, **17**, 89–99.
51. Alexandrie,A.K. *et al*. (2004) Influence of CYP1A1, GSTM1, GSTT1, and NQO1 genotypes and cumulative smoking dose on lung cancer risk in a Swedish population. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 908–914.
52. Vineis,P. *et al*. (2004) CYP1A1, GSTM1 and GSTT1 polymorphisms and lung cancer: a pooled analysis of gene-gene interactions. *Biomarkers*, **9**, 298–305.
53. Raimondi,S. *et al*. (2005) Metabolic gene polymorphisms and lung cancer risk in non-smokers: an update of the GSEC study. *Mutat. Res.*, **592**, 45–57.
54. Larsen,J.E. *et al*. (2006) CYP1A1 Ile462Val and MPO G-463A interact to increase risk of adenocarcinoma but not squamous cell carcinoma of the lung. *Carcinogenesis*, **27**, 525–532.
55. Liu,G. *et al*. (2004) MPO and SOD2 polymorphisms, gender, and the risk of non-small cell lung carcinoma. *Cancer Lett.*, **214**, 69–79.
56. Zhou,W. *et al*. (2002) Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 15–21.
57. Spitz,M.R. *et al*. (2003) Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 689–698.
58. Mohrenweiser,H.W. *et al*. (2002) Identification of 127 amino acid substitution variants in screening 37 DNA repair genes in humans. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1054–1064.
59. Shen,M.R. *et al*. (1998) Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res.*, **58**, 604–608.
60. Hung,R.J. *et al*. (2005) Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am. J. Epidemiol.*, **162**, 925–942.
61. Goode,E.L. *et al*. (2002) Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1513–1530.
62. Hung,R.J. *et al*. (2005) Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. *J. Natl Cancer Inst.*, **97**, 567–576.
63. Zhou,W. *et al*. (2003) Polymorphisms in the DNA repair genes XRCC1 and ERCC2, smoking, and lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 359–365.
64. Yoon,S.M. *et al*. (2005) The polymorphism and haplotypes of XRCC1 and survival of non-small-cell lung cancer after radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, **63**, 885–891.
65. Hu,Z. *et al*. (2004) DNA repair gene XPD polymorphism and lung cancer risk: a meta-analysis. *Lung Cancer*, **46**, 1–10.
66. Benhamou,S. *et al*. (2005) ERCC2/XPD gene polymorphisms and lung cancer: a HuGE review. *Am. J. Epidemiol.*, **161**, 1–14.
67. Booton,R. *et al*. (2006) Xeroderma pigmentosum group D haplotype predicts for response, survival, and toxicity after platinum-based chemotherapy in advanced nonsmall cell lung cancer. *Cancer*, **106**, 2421–2427.
68. Ryu,J.S. *et al*. (2004) Association between polymorphisms of ERCC1 and XPD and survival in non-small-cell lung cancer patients treated with cisplatin combination chemotherapy. *Lung Cancer*, **44**, 311–316.
69. Zienolddiny,S. *et al*. (2006) Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis*, **27**, 560–567.
70. Olaussen,K.A. *et al*. (2006) DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N. Engl. J. Med.*, **355**, 983–991.
71. Zhang,X. *et al*. (2005) Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. *Cancer Res.*, **65**, 722–726.
72. Coussens,L.M. *et al*. (2002) Inflammation and cancer. *Nature*, **420**, 860–867.
73. Skillrud,D.M. *et al*. (1986) Higher risk of lung cancer in chronic obstructive pulmonary disease. A prospective, matched, controlled study. *Ann. Intern. Med.*, **105**, 503–507.
74. Tockman,M.S. *et al*. (1987) Airways obstruction and the risk for lung cancer. *Ann. Intern. Med.*, **106**, 512–518.
75. Bernert,H. *et al*. (2003) Tnfa and Il-10 deficiencies have contrasting effects on lung tumor susceptibility: gender-dependent modulation of IL-10 haploinsufficiency. *Mol. Carcinog.*, **38**, 117–123.
76. Langowski,J.L. *et al*. (2006) IL-23 promotes tumour incidence and growth. *Nature*, **442**, 461–465.
77. Zienolddiny,S. *et al*. (2004) Polymorphisms of the interleukin-1 beta gene are associated with increased risk of non-small cell lung cancer. *Int. J. Cancer*, **109**, 353–356.
78. Lind,H. *et al*. (2005) Interleukin 1 receptor antagonist gene polymorphism and risk of lung cancer: a possible interaction with polymorphisms in the interleukin 1 beta gene. *Lung Cancer*, **50**, 285–290.
79. Campa,D. *et al*. (2004) Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. *Carcinogenesis*, **25**, 229–235.
80. Nagase,H. *et al*. (1999) Matrix metalloproteinases. *J. Biol. Chem.*, **274**, 21491–21494.
81. Overall,C.M. *et al*. (2006) Towards third generation matrix metalloproteinase inhibitors for cancer therapy. *Br. J. Cancer*, **94**, 941–946.
82. Su,L. *et al*. (2006) Genotypes and haplotypes of matrix metalloproteinase 1, 3 and 12 genes and the risk of lung cancer. *Carcinogenesis*, **27**, 1024–1029.
83. Su,L. *et al*. (2005) Matrix metalloproteinase-1 promoter polymorphism and lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 567–570.

84. Yu,C. *et al*. (2002) Correlation between a single nucleotide polymorphism in the matrix metalloproteinase-2 promoter and risk of lung cancer. *Cancer Res.*, **62**, 6430–6433.
85. Heist,R.S. *et al*. (2006) Matrix metalloproteinase polymorphisms and survival in stage I non-small cell lung cancer. *Clin. Cancer Res.*, **12**, 5448–5453.
86. Khuri,F.R. *et al*. (2001) Cyclooxygenase-2 overexpression is a marker of poor prognosis in stage I non-small cell lung cancer. *Clin. Cancer Res.*, **7**, 861–867.
87. Wolff,H. *et al*. (1998) Expression of cyclooxygenase-2 in human lung carcinoma. *Cancer Res.*, **58**, 4997–5001.
88. Laga,A.C. *et al*. (2005) Prognostic significance of cyclooxygenase 2 expression in 259 cases of non-small cell lung cancer. *Arch. Pathol. Lab. Med.*, **129**, 1113–1117.
89. Harris,R.E. *et al*. (2002) Chemoprevention of lung cancer by non-steroidal anti-inflammatory drugs among cigarette smokers. *Oncol. Rep.*, **9**, 693–695.
90. Campa,D. *et al*. (2005) Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. *Carcinogenesis*, **26**, 1157.
91. Lynch,T.Jr. *et al*. (2005) Optimizing chemotherapy and targeted agent combinations in NSCLC. *Lung Cancer*, **50**, (Suppl. 2), S25–S32.
92. Zhang,X. *et al*. (2006) Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Hum. Mutat.*, **27**, 110–117.
93. Schabath,M.B. *et al*. (2006) Combined effects of the p53 and p73 polymorphisms on lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 158–161.
94. Zhang,X. *et al*. (2005) Functional polymorphisms in cell death pathway genes FAS and FASL contribute to risk of lung cancer. *J. Med. Genet.*, **42**, 479–484.
95. Burke,L. *et al*. (2005) Prognostic implications of molecular and immunohistochemical profiles of the Rb and p53 cell cycle regulatory pathways in primary non-small cell lung carcinoma. *Clin. Cancer Res.*, **11**, 232–241.
96. Huncharek,M. *et al*. (2000) Prognostic significance of p53 mutations in non-small cell lung cancer: a meta-analysis of 829 cases from eight published studies. *Cancer Lett.*, **153**, 219–226.
97. Mitsudomi,T. *et al*. (2000) Prognostic significance of p53 alterations in patients with non-small cell lung cancer: a meta-analysis. *Clin. Cancer Res.*, **6**, 4055–4063.
98. Steels,E. *et al*. (2001) Role of p53 as a prognostic factor for survival in lung cancer: a systematic review of the literature with a meta-analysis. *Eur. Respir. J.*, **18**, 705–719.
99. Nelson,H.H. *et al*. (2005) TP53 mutation, allelism and survival in non-small cell lung cancer. *Carcinogenesis*, **26**, 1770–1773.
100. Kato,S. *et al*. (1995) Human lung carcinogen-DNA adduct levels mediated by genetic polymorphisms *in vivo*. *J. Natl Cancer Inst.*, **87**, 902–907.
101. Ryberg,D. *et al*. (1997) Genotypes of glutathione transferase M1 and P1 and their significance for lung DNA adduct levels and cancer risk. *Carcinogenesis*, **18**, 1285–1289.
102. Rojas,M. *et al*. (1998) High benzo(a)pyrene diol-epoxide DNA adduct levels in lung and blood cells from individuals with combined CYP1A1 MspI/Msp-GSTM1*0/*0 genotypes. *Pharmacogenetics*, **8**, 109–118.
103. Hou,S.M. *et al*. (2001) Differential interactions between GSTM1 and NAT2 genotypes on aromatic DNA adduct level and HPRT mutant frequency in lung cancer patients and population controls. *Cancer Epidemiol. Biomarkers Prev.*, **10**, 133–140.
104. Hou,S.M. *et al*. (1995) Relationship between hprt mutant frequency, aromatic DNA adducts and genotypes for GSTM1 and NAT2 in bus maintenance workers. *Carcinogenesis*, **16**, 1913–1917.
105. Ichiba,M. *et al*. (1994) Aromatic DNA adducts, micronuclei and genetic polymorphism for CYP1A1 and GST1 in chimney sweeps. *Carcinogenesis*, **15**, 1347–1352.
106. Peluso,M. *et al*. (2004) Comparison of DNA adduct levels in nasal mucosa, lymphocytes and bronchial mucosa of cigarette smokers and interaction with metabolic gene polymorphisms. *Carcinogenesis*, **25**, 2459–2465.
107. Rothman,N. *et al*. (1995) The impact of glutathione s-transferase M1 and cytochrome P450 1A1 genotypes on white-blood-cell polycyclic aromatic hydrocarbon-DNA adduct levels in humans. *Mol. Carcinog.*, **14**, 63–68.
108. Santella,R.M. *et al*. (1995) Polycyclic aromatic hydrocarbon-DNA and protein adducts in coal tar treated patients and controls and their relationship to glutathione S-transferase genotype. *Mutat. Res.*, **334**, 117–124.
109. Schoket,B. *et al*. (1998) Smoking-associated bulky DNA adducts in bronchial tissue related to CYP1A1 MspI and GSTM1 genotypes in lung patients. *Carcinogenesis*, **19**, 841–846.
110. Hou,S.M. *et al*. (2002) The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. *Carcinogenesis*, **23**, 599–603.
111. Palli,D. *et al*. (2001) DNA adduct levels and DNA repair polymorphisms in traffic-exposed workers and a general population sample. *Int. J. Cancer*, **94**, 121–127.
112. Matullo,G. *et al*. (2001) XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. *Carcinogenesis*, **22**, 1437–1445.
113. Mechanic,L.E. *et al*. (2005) Polymorphisms in XPD and TP53 and mutation in human lung cancer. *Carcinogenesis*, **26**, 597–604.
114. Belinsky,S.A. (2005) Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. *Carcinogenesis*, **26**, 1481–1487.
115. Zochbauer-Muller,S. *et al*. (2002) Aberrant DNA methylation in lung cancer: biological and clinical implications. *Oncologist*, **7**, 451–457.
116. Belinsky,S.A. *et al*. (2005) Gene promoter methylation in plasma and sputum increases with lung cancer risk. *Clin. Cancer Res.*, **11**, 6505–6511.
117. Fujiwara,K. *et al*. (2005) Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. *Clin. Cancer Res.*, **11**, 1219–1225.
118. Pulling,L.C. *et al*. (2003) Promoter hypermethylation of the O6-methylguanine-DNA methyltransferase gene: more common in lung adenocarcinomas from never-smokers than smokers and associated with tumor progression. *Cancer Res.*, **63**, 4842–4848.
119. Toyooka,S. *et al*. (2003) Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. *Int. J. Cancer*, **103**, 153–160.
120. Belinsky,S.A. (2004) Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat. Rev. Cancer*, **4**, 707–717.
121. Bowman,R.V. *et al*. (2006) Epigenetics of lung cancer. *Respirology*, **11**, 355–365.
122. Hung,R.J. *et al*. (2005) Perspectives on the molecular epidemiology of aerodigestive tract cancers. *Mutat. Res.*, **592**, 102–118.
123. Toyooka,S. *et al*. (2004) The relationship between aberrant methylation and survival in non-small-cell lung cancers. *Br. J. Cancer*, **91**, 771–774.
124. Kim,J.S. *et al*. (2006) Cohypermethylation of p16 and FHIT promoters as a prognostic factor of recurrence in surgically resected stage I non-small cell lung cancer. *Cancer Res.*, **66**, 4049–4054.
125. Kim,Y.T. *et al*. (2005) Prognostic implication of aberrant promoter hypermethylation of CpG islands in adenocarcinoma of the lung. *J. Thorac. Cardiovasc. Surg.*, **130**, 1378.
126. Safar,A.M. *et al*. (2005) Methylation profiling of archived non-small cell lung cancer: a promising prognostic system. *Clin. Cancer Res.*, **11**, 4400–4405.
127. Gilliland,F.D. *et al*. (2002) Glutathione S-transferase P1 and NADPH quinone oxidoreductase polymorphisms are associated with aberrant promoter methylation of P16(INK4a) and O(6)-methylguanine-DNA methyltransferase in sputum. *Cancer Res.*, **62**, 2248–2252.
128. Chan,E.C. *et al*. (2005) Polymorphisms of the GSTM1, GSTP1, MPO, XRCC1, and NQO1 genes in Chinese patients with non-small cell lung cancers: relationship with aberrant promoter methylation of the CDKN2A and RARB genes. *Cancer Genet. Cytogenet.*, **162**, 10–20.
129. Institute, U.o.P.C. (2006) GSEC: International Collaborative Study on Genetic Susceptibility to Environmental Carcinogens: An International Pooled Data Analysis on Metabolic Susceptibility Gene Polymorphisms. University of Pittsburgh Cancer Institute (UPCI), Pittsburgh, PA, vol. 2006.
130. Blettner,M. *et al*. (1999) Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int. J. Epidemiol.*, **28**, 1–9.
131. Taioli,E. *et al*. (2002) Methodological issues in pooled analysis of biomarker studies. *Mutat. Res.*, **512**, 85–92.
132. The international HapMap project. *Nature*, **426**, 789–796.
133. Montpetit,A. *et al*. (2006) An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. *PLoS Genet.*, **2**, e27.
134. de Bakker,P.I. *et al*. (2006) Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations. *Pac. Symp. Biocomput.*, pp. 478–486.
135. Ribas,G. *et al*. (2006) Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Hum. Genet.*, **118**, 669–679.
136. Willer,C.J. *et al*. (2006) Tag SNP selection for Finnish individuals based on the CEPH Utah HapMap database. *Genet. Epidemiol.*, **30**, 180–190.
137. Tantoso,E. *et al*. (2006) How well do HapMap SNPs capture the untyped SNPs? *BMC Genomics*, **7**, 238.

138. Risch,N. *et al*. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
139. Houlston,R.S. *et al*. (2004) The search for low-penetrance cancer susceptibility alleles. *Oncogene*, **23**, 6471–6476.
140. McKeigue,P.M. (2005) Prospects for admixture mapping of complex traits. *Am. J. Hum. Genet.*, **76**, 1–7.
141. David-Beabes,G.L. *et al*. (2001) Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians. *Lung Cancer*, **34**, 333–339.
142. Cote,M.L. *et al*. (2005) Combinations of glutathione S-transferase genotypes and risk of early-onset lung cancer in Caucasians and African Americans: a population-based study. *Carcinogenesis*, **26**, 811–819.
143. Ford,J.G. *et al*. (2000) Glutathione S-transferase M1 polymorphism and lung cancer risk in African-Americans. *Carcinogenesis*, **21**, 1971–1975.
144. Cote,M.L. *et al*. (2005) Risk of lung cancer among white and black relatives of individuals with early-onset lung cancer. *JAMA*, **293**, 3036–3042.
145. Network,T.E.D.R. (2000) Translational research to identify early cancer and cancer risk. National Cancer Institute Division of Cancer Prevention, Bethesda, MD.
146. Normolle,D. *et al*. (2005) Design of early validation trials of biomarkers. *Cancer Inform.*, **1**, 25–32.
147. Benhamou,S. *et al*. (2002) ERCC2/XPD gene polymorphisms and cancer risk. *Mutagenesis*, **17**, 463–469.

Received August 2, 2006; revised December 15, 2006; accepted December 17, 2006