Variants in the *GH–IGF* axis confer susceptibility to lung cancer

Matthew F. Rudd,[1] Emily L. Webb,[1] Athena Matakidou,[1] Gabrielle S. Sellick,[1] Richard D. Williams,[2] Helen Bridle,[3] Tim Eisen,[3,4] Richard S. Houlston,[1,4,5] and the GELCAPS Consortium[6]

[1]*Section of Cancer Genetics,* [2]*Section of Paediatrics, and* [3]*Section of Medicine, Institute of Cancer Research, Sutton, Surrey SM2 5NG, United Kingdom*

We conducted a large-scale genome-wide association study in UK Caucasians to identify susceptibility alleles for lung cancer, analyzing 1529 cases and 2707 controls. To increase the likelihood of identifying disease-causing alleles, we genotyped 1476 nonsynonymous single nucleotide polymorphisms (nsSNPs) in 871 candidate cancer genes, biasing SNP selection toward those predicted to be deleterious. Statistically significant associations were identified for 64 nsSNPs, generating a genome-wide significance level of $P = 0.002$. Eleven of the 64 SNPs mapped to genes encoding pivotal components of the growth hormone/insulin-like growth factor (*GH–IGF*) pathway, including *CAMKK1* E375G (OR = 1.37, $P = 5.4 \times 10^{-5}$), *AKAP9* M4631 (OR = 1.32, $P = 1.0 \times 10^{-4}$) and *GHR* P495T (OR = 12.98, $P = 0.0019$). Significant associations were also detected for SNPs within genes in the DNA damage-response pathway, including *BRCA2* K3326X (OR = 1.72, $P = 0.0075$) and *XRCC4* I137T (OR = 1.31, $P = 0.0205$). Our study provides evidence that inherited predisposition to lung cancer is in part mediated through low-penetrance alleles and specifically identifies variants in *GH–IGF* and DNA damage-response pathways with risk of lung cancer.

[Supplemental material is available online at www.genome.org.]

Lung cancer is the most common cancer in the world and represents a major public health problem, accounting for ~1.2 million cancer-related deaths worldwide each year (Parkin et al. 2005). Tobacco smoking is acknowledged to be the major risk factor for lung cancer, contributing to a 10-fold increase in risk in long-term smokers compared with nonsmokers (Doll and Peto 1981). Other environmental risk factors include exposure to radiation, asbestos, heavy metals, polycyclic aromatic hydrocarbons, and chloromethyl ethers (IARC 1986).

Lung cancer is frequently cited as a malignancy solely attributable to environmental exposure. However, it has long been postulated that individuals may differ in their susceptibility and there is increasing evidence from epidemiological studies for a familial risk (Matakidou et al. 2005). Direct evidence for a genetic predisposition is provided by the increased risk of lung cancer associated with a number of rare Mendelian cancer syndromes, such as carriers of constitutional tumor protein p53 (*TP53*) (Hwang et al. 2003) and retinoblastoma (Sanders et al. 1989) gene mutations, as well as in patients with Bloom's (Takemiya et al. 1987) and Werner's syndromes (Yamanaka et al. 1997).

The genetic basis of inherited susceptibility to lung cancer outside the context of the rare Mendelian cancer predisposition syndromes is at present undefined, but a model in which dominantly acting, high-risk alleles account for all of the excess familial risk seems unlikely. An alternative hypothesis about the allelic architecture of lung cancer susceptibility proposes that most of the genetic risk is caused by low-penetrance alleles. This hypothesis implies that testing for allelic association should be a powerful strategy for identifying lung cancer predisposition alleles.

We sought to identify novel low-penetrance susceptibility alleles to lung cancer by genotyping SNPs across 871 genes with relevance to cancer biology. To increase the likelihood of identifying disease-causing alleles, we biased selection of nsSNPs to those likely to have functionally deleterious consequences. Genotyping 1529 lung cancer cases and 2707 controls from the UK population across 1476 nsSNPs provided strong evidence that low-penetrance alleles in genes involved in the hormone/insulin-like growth factor (*GH-IGF*) and DNA damage-response pathways are associated with lung cancer susceptibility

Results

Genotypes were obtained for 1526 cases (99.8%) and 2695 controls (99.6%). Of the 1476 SNPs submitted for analysis, 1221 SNPs had sample call rates >95%. Of these, 180 were fixed, leaving 1041 SNPs for which genotype data were informative (Supplemental Table 1). Implementing the genomic control method indicated no evidence of population stratification in our data as a cause of false-positive results, as the 95% confidence interval for the stratification parameter $\hat{\lambda}$(0.92–1.31) encompassed unity. As deviates from Hardy-Weinberg equilibrium followed the expected distribution, we concluded that genotyping error is unlikely to have impacted on the statistics generated.

Significant associations with risk of lung cancer were identified for 64 of 1041 nsSNPs at the 5% level. The overrepresentation of associations between SNPs and lung cancer risk was confirmed by a joint analysis of their combined effect using the set-association approach (smallest global significance level of $P = 4.2 \times 10^{-4}$). After further adjustment for the number of

[4]**These two authors contributed equally to this work.**
[5]**Corresponding author.**
E-mail Richard.Houlston@icr.ac.uk; fax 4-20-8722-4359.
[6]**List of GELCAPS Consortium collaborators available on request.**
Article is online at http://www.genome.org/cgi/doi/10.1101/gr.5120106.

terms in the set being a priori, unknown, the genome-wide significance was $P = 0.002$.

Two of the 64 SNPs identified through the set association procedure, rs2602141 (K1136Q) and rs560191 (D353E), map to the tumor protein p53 binding protein 1 (*TP53BP1*) and are in strong linkage disequilibrium (LD). A further group of three SNPs in the MHC region spaced within 100 kb; rs1052486 (S625P) in HLA-B-associated transcript 3 (*BAT3*), rs3130618 (R41L) in HLA-B-associated transcript 4 (*BAT4*), and rs16900023 (P786S) in mutS homolog 5 (*MSH5*) also formed a cluster of high LD. Although the permutation procedure implemented in the set-association strategy allows for such substructure in the data when estimating significance levels, it may not be desirable to include highly correlated SNPs in the analysis. A total of 262 SNPs displayed high LD with an adjacent SNP. High LD was found to occur primarily within the same gene, but there were 80 instances where strong LD was observed between SNPs in different genes. We repeated our analysis by omitting markers in LD, retaining one SNP per LD set on the basis of maximum GenCall score or call rate, yielding almost identical sum statistics ($P = 0.005$) with inclusion of 70 SNPs.

Sixty-seven SNPs displayed significant association at the 5% level with familial lung cancer, but only 52 when the analysis was restricted to sporadic cases. After permutation, the overall significance level attained from the set-association analysis for the familial cases was $P = 0.015$ compared with $P = 0.076$ for sporadic cases. Familial cases contributed significantly to overall study findings with 13 SNPs contributing to the 20 associated at the 1% level in the overall data set (Table 1). Stratification of cases by cancer histology (small cell and non-small cell, global *P*-values 0.11 and 0.06, respectively), age at diagnosis (<60 and ≥60; global *P*-values 0.17 and 0.18, respectively) and sex (male and female; global *P*-values 0.19 and 0.08, respectively) did not impact significantly on study findings. Furthermore, limiting our analysis to the 93.7% of cases who were smokers indicated that there was no evidence of confounding due to smoking.

The SNP showing the most significant allelic association with lung cancer was rs1052486 (S625P) in *BAT3*, a nuclear protein implicated in the control of apoptosis, with strongest association under a recessive model ($OR_R = 0.69$, 95% CI: 0.59–0.82, $P_R = 8.3 \times 10^{-6}$) (Table 1). Two additional SNPs, rs7214723 (E375G) in calcium/calmodulin-dependent protein kinase kinase 1 α (*CAMKK1*), belonging to the Serine/Threonine protein kinase family ($OR_R = 1.37$, 95% CI: 1.17–1.59, $P_R = 5.4 \times 10^{-5}$) and rs6964587 (M463I) in A kinase anchor protein 9 (*AKAP9*), a key component of signal transduction ($OR_D = 1.32$, 95% CI: 1.15–1.52, $P_D = 7.6 \times 10^{-5}$), also showed highly significant nominal association under recessive and dominant models, respectively. Empirical limits for genome-wide significance for individual T_A, T_D, and T_R statistics were established at 16.12, 16.23, and 15.66, respectively. Hence, *BAT3* S625P and *CAMKK1* E375G were both significantly associated with lung cancer with adjusted *P*-values of 0.006 and 0.036, respectively, with *AKAP9* M463I showing borderline significance with adjusted $P = 0.066$.

Of the 64 SNPs identified, two SNPs have been documented to be functional, i.e., K3326X in breast cancer 2 early onset (*BRCA2*) and N700S in thrombospondin 1 (*THBS1*), and a further 37 SNPs are predicted in silico to deleteriously impact on the expressed proteins (Table 2).

Through interrogation of the Pathway Assist program (Stratagene), 11 of the 64 SNPs associated with risk of lung cancer were located within individual genes encoding pivotal components of the extended *GH-IGF* pathway, including *CAMKK1* E375G and *AKAP9* M463I (both of which were globally significant), growth hormone receptor (*GHR*) P495T ($OR_D = 12.98$, $P_D = 0.0019$), A kinase anchor protein 10 (*AKAP10*) R249H ($OR_R = 1.25$, $P_R = 0.0085$), and insulin-like growth-factor binding protein 5 (*IGFBP5*) R138W ($OR_D = 1.29$, $P_D = 0.027$) (Table 1). A further five SNPs were located in genes directly involved in the DNA damage-response pathway, including the functional *BRCA2* SNP K3326X ($OR_D = 1.72$, $P_D = 0.0075$), X-ray repair complementing defective repair in Chinese hamster cells 4 (*XRCC4*) I134T ($OR_D = 1.31$, $P_D = 0.0205$), mutS homolog 5 (*MSH5*) P786S ($OR_D = 0.64$, $P_D = 0.0228$), mutS homolog 4 (*MSH4*) S914N ($OR_D = 1.27$, $P_D = 0.0461$), and *BRCA1*-associated RING domain 1 (*BARD1*) R658C ($OR_D = 1.59$, $P_D = 0.0329$) (Table 1).

Haplotype frequencies defined by the two sets of SNPs displaying high LD, *TP53BP1* K1136Q and D353E, and *BAT3* S625P, *BAT4* R41L, and *MSH5* P786S were significantly different in cases and controls (adjusted *P*-values, 0.01 and 0.01, respectively, after permutation testing).

We examined for potential interactive effects between the 64 SNPs significantly associated with lung cancer risk ($P_A < 0.05$) by fitting full logistic regression models for each pair, generating 2016 models, and comparing these with the main effects model. Ninety-six pairs of SNPs showed nominally significant interaction at the 5% level. The largest interactive effect identified was between 1-aminocyclopropane-1-carboxylate synthase (*PHACS*) P421L and toll-like receptor 1 (*TLR1*) R80T ($P = 3.3 \times 10^{-4}$), albeit nonsignificant after correction for multiple testing.

Discussion

To date, the only evidence for a major locus for lung cancer susceptibility is provided by the linkage scan conducted by Bailey-Wilson et al. (2004), which reported linkage of the disease to chromosome 6q23–25, and a model based on involvement of multiple low-penetrance alleles is eminently plausible.

Previous association studies aimed at identifying low-penetrance alleles for lung cancer susceptibility have evaluated a restricted number of polymorphisms, primarily in genes implicated in the metabolism of tobacco-associated carcinogens and protection of DNA from carcinogen-induced damage. To identify novel lung cancer susceptibility alleles, we extended our search to include genes with relevance to cancer biology, evaluating only nsSNPs that have a higher probability of being directly causal. We acknowledge that the loci considered as candidates will be based on current preconceptions of cancer biology, and it is likely that other genes may influence tumor development. The number of candidate loci will inevitably increase with advances in cancer biology.

The number of nsSNPs that displayed significant association with lung cancer risk was greater than that expected, supporting the tenet that polymorphic variation contributes to lung cancer susceptibility. This assertion is supported by the fact that associations were stronger when the analysis was restricted to those cases with a family history of lung cancer. We cannot exclude the possibility that some of the associations detected are a consequence of LD with causal mutations. It is noteworthy that the SNPs in *BAT3*, *BAT4*, and *MSH5*, which were all associated with lung cancer risk, were in strong LD.

Of the 64 SNPs found to be associated with lung cancer risk, several reside in genes involved in either apoptosis (*BARD1* and

Table 1. SNPs showing significant allelic association with lung cancer

				Allelic statistic		Dominant/Recessive statistics	
SNP	Gene[a]	Substitution	MAF[b]	OR (95% CI)	P_A	$OR_{D/R}$ (95% CI)[c]	$P_{D/R}$[c]
rs1052486[d]	*BAT3*	S625P	0.483	0.84 (0.77,0.92)	0.0002	0.69 (0.59,0.82)^R	$8.3x10^{-6}$ ^{R}
rs7214723[d]	*CAMKK1*	E375G	0.446	1.18 (1.08,1.29)	0.0003	1.37 (1.17,1.59)^R	$5.4x10^{-5}$ ^{R}
rs560191[d]	*TP53BP1*	D353E	0.305	0.85 (0.77,0.93)	0.0009	0.84 (0.74,0.95)^D	0.0050^D
rs3130618[d]	*BAT4*	R41L	0.190	1.20 (1.07,1.34)	0.0013	1.26 (1.11,1.44)^D	0.0005^D
rs2602141[d]	*TP53BP1*	K1136Q	0.304	0.85 (0.77,0.94)	0.0014	0.84 (0.74,0.95)^D	0.0063^D
rs6964587[d]	*AKAP9*	M463I	0.383	1.16 (1.06,1.28)	0.0016	1.32 (1.15,1.52)^D	0.0001^D
rs2229742	*NRIP1*	R448G	0.101	1.25 (1.08,1.43)	0.0021	1.24 (1.07,1.45)^D	0.0052^D
rs2660744	*PPAT*	Q488X	0.155	0.82 (0.72,0.93)	0.0024	0.80 (0.69,0.93)^D	0.0026^D
rs3206824	*DKK3*	R335G	0.231	1.16 (1.05,1.29)	0.0046	1.21 (1.07,1.38)^D	0.0028^D
rs6183[d]	*GHR*	P495T	0.001	12.98 (1.77, ∞)	0.0047	12.98 (1.77, ∞)^D	0.0019^D
rs1129923[d]	*DUSP23*	G131S	0.097	0.80 (0.68,0.93)	0.0050	0.79 (0.67,0.94)^D	0.0069^D
rs11571833[d]	*BRCA2*	K3326X	0.009	1.74 (1.17,2.59)	0.0054	1.72 (1.15,2.57)^D	0.0075^D
rs2242089	*PYCRL*	V105M	0.183	0.84 (0.75,0.95)	0.0054	0.81 (0.71,0.93)^D	0.0028^D
rs1738023[d]	*AKR7A3*	N215D	0.164	1.18 (1.05,1.32)	0.0063	1.19 (1.04,1.37)^D	0.0098^D
rs11569705	*SULT1E1*	D22Y	0.003	0.10 (0.01,0.78)	0.0068	0.10 (0.01,0.78)^D	0.0068^D
rs2295778[d]	*HIF1AN*	P41A	0.263	1.15 (1.04,1.26)	0.0070	1.34 (1.07,1.69)^R	0.0112^R
rs2306022	*ITGA11*	V433M	0.094	0.80 (0.68,0.94)	0.0072	0.79 (0.66,0.94)^D	0.0065^D
rs10115703	*CER1*	R19W	0.079	0.79 (0.66,0.94)	0.0088	0.76 (0.63,0.92)^D	0.0041^D
rs2108978[d]	*AKAP10*	R249H	0.393	1.13 (1.03,1.23)	0.0092	1.25 (1.06,1.48)^R	0.0085^R
rs2725362[d]	*WRN*	L1074F	0.473	0.89 (0.81,0.97)	0.0100	0.81 (0.69,0.94)^R	0.0069^R
rs3732401[d]	*GTF2E1*	P366S	0.043	0.73 (0.57,0.93)	0.0112	0.72 (0.57,0.93)^D	0.0107^D
rs4371716[d]	*CDH12*	V68M	0.242	1.14 (1.03,1.26)	0.0122	1.63 (1.27,2.09)^R	0.0001^R
rs8065506[d]	*ZNF624*	K135N	0.265	1.13 (1.03,1.25)	0.0133	1.14 (1.01,1.30)^D	0.0355^D
rs970547[d]	*COL12A1*	G1894S	0.223	0.87 (0.78,0.97)	0.0141	0.67 (0.49,0.93)^R	0.0173^R
rs2032729	*ZNF24*	N220S	0.076	1.22 (1.04,1.43)	0.0143	1.24 (1.05,1.47)^D	0.0124^D
rs2243639	*SFTPD*	T180A	0.408	0.89 (0.82,0.98)	0.0151	0.87 (0.76,0.99)^D	0.0330^D
rs17632786[d]	*THBS1*	N700S	0.133	0.85 (0.74,0.97)	0.0163	0.81 (0.70,0.94)^D	0.0056^D
rs10787428	*GPAM*	E131G	0.396	0.89 (0.82,0.98)	0.0168	0.87 (0.77,0.99)^D	0.0350^D
rs17184326	*POP1*	K522N	0.128	1.17 (1.03,1.33)	0.0174	1.18 (1.02,1.36)^D	0.0257^D
rs11575194[d]	*IGFBP5*	R138W	0.038	1.29 (1.04,1.61)	0.0183	1.29 (1.03,1.61)^D	0.0270^D
rs363504	*GRIK1*	L902S	0.051	0.77 (0.62,0.96)	0.0194	0.15 (0.02,1.13)^R	0.0324^R
rs28360135	*XRCC4*	I134T	0.035	1.31 (1.04,1.64)	0.0202	1.31 (1.04,1.66)^D	0.0205^D
rs1051740[d]	*EPHX1*	Y113H	0.286	1.12 (1.02,1.24)	0.0204	1.15 (1.01,1.30)^D	0.0297^D
rs16900023[d]	*MSH5*	P786S	0.018	0.64 (0.43,0.94)	0.0206	0.64 (0.43,0.94)^D	0.0228^D
rs1043261	*IL17RB*	Q484X	0.084	0.82 (0.69,0.97)	0.0222	0.82 (0.69,0.99)^D	0.0334^D
rs2295000[d]	*DATF1*	S535L	0.218	0.88 (0.79,0.98)	0.0224	0.86 (0.75,0.98)^D	0.0202^D
rs1019670	*MS4A6A*	N150I	0.393	0.90 (0.82,0.99)	0.0243	0.88 (0.77,1.00)^D	0.0455^D
rs7998427	*SETDB2*	E117G	0.324	0.90 (0.81,0.99)	0.0248	0.87 (0.77,0.99)^D	0.0367^D
rs9262138	*DHX16*	D566G	0.061	0.80 (0.65,0.97)	0.0249	0.79 (0.64,0.97)^D	0.0230^D
rs5745549	*MSH4*	S914N	0.033	1.30 (1.03,1.63)	0.0256	1.27 (1.00,1.61)^D	0.0461^D
rs3107275	*PHACS*	P421L	0.406	0.90 (0.82,0.99)	0.0272	0.85 (0.71,1.01)^R	0.0609^R
rs8069344	*GUCY2D*	L782H	0.135	0.86 (0.75,0.98)	0.0277	0.82 (0.70,0.95)^D	0.0078^D
rs5388	*GH1*	V136I	0.012	0.59 (0.36,0.95)	0.0287	0.56 (0.34,0.91)^D	0.0185^D
rs1800974	*ITGA7*	R651H	0.488	1.10 (1.01,1.21)	0.0307	1.14 (0.98,1.31)^R	0.0803^R
rs1820128	*ZNF600*	C209R	0.143	0.87 (0.76,0.99)	0.0325	0.85 (0.74,0.98)^D	0.0294^D
rs17246389	*SERPINI2*	L6V	0.266	0.89 (0.81,0.99)	0.0328	0.86 (0.76,0.98)^D	0.0227^D
rs17337252	*RB1CC1*	M234T	0.491	1.10 (1.01,1.20)	0.0329	1.15 (1.00,1.33)^R	0.0584^R
rs3758938	*TBX10*	K101T	0.312	0.90 (0.82,0.99)	0.0338	0.88 (0.78,1.00)^D	0.0551^D
rs1800076	*CFTR*	R75Q	0.037	1.27 (1.02,1.58)	0.0339	1.27 (1.01,1.59)^D	0.0412^D
rs2230674[d]	*ATF1*	P191A	0.041	0.77 (0.60,0.98)	0.0348	0.77 (0.60,0.99)^D	0.0425^D
rs2274750	*TNC*	A1781T	0.025	1.32 (1.01,1.71)	0.0391	1.32 (1.01,1.73)^D	0.0391^D
rs933135	*PLCD1*	R257H	0.013	1.45 (1.02,2.06)	0.0391	1.48 (1.03,2.11)^D	0.0312^D
rs5743611	*TLR1*	R80T	0.087	0.84 (0.71,0.99)	0.0394	0.84 (0.70,1.00)^D	0.0525^D
rs1211554	*HUS1B*	D268Y	0.089	0.84 (0.72,0.99)	0.0415	0.85 (0.72,1.01)^D	0.0694^D
rs4647932	*FGFRL1*	P464L	0.061	1.20 (1.01,1.43)	0.0415	1.19 (0.99,1.44)^D	0.0652^D
rs17356233	*CHD1L*	H350Q	0.247	0.90 (0.81,1.00)	0.0417	0.86 (0.76,0.98)^D	0.0220^D
rs3738888[d]	*BARD1*	R658C	0.009	1.55 (1.01,2.36)	0.0423	1.59 (1.03,2.44)^D	0.0329^D
rs12500797	*PTPN13*	E1606K	0.107	1.15 (1.00,1.32)	0.0443	1.18 (1.01,1.37)^D	0.0364^D
rs4988492	*GHRH*	L75F	0.013	1.44 (1.00,2.05)	0.0470	1.44 (1.01,2.07)^D	0.0453^D
rs2230339	*GPR68*	R63Q	0.001	6.80 (0.73, ∞)	0.0472	6.80 (0.73, ∞)^D	0.0213^D
rs2229424	*FASN*	R1694H	0.001	6.80 (0.73, ∞)	0.0474	6.79 (0.73, ∞)^D	0.0214^D
rs4791641	*PFAS*	P367L	0.497	0.91 (0.84,1.00)	0.0475	0.87 (0.75,1.01)^R	0.0668^R
rs11652709	*EPX*	Q122H	0.320	1.10 (1.00,1.21)	0.0477	1.13 (0.99,1.28)^D	0.0668^D
rs1801690	*APOH*	W335S	0.056	0.81 (0.66,1.00)	0.0480	0.81 (0.66,1.00)^D	0.0549^D

[a]NCBI Entrez Gene.
[b]Minor allele frequency (MAF) in cases.
[c]Most significant association under a dominant (D) or recessive (R) model.
[d]Associated at significance level 5% when analysis restricted to familial cases.

Table 2. Description and predicted functionality of nsSNPs showing significant association with lung cancer risk

SNP	Substitution	Predicted Functionality[a]	Gene[b]	Gene Description	Gene Ontology[c]	OMIM[d]
rs1052486	S625P	Possibly damaging	*BAT3*	HLA-B associated transcript 3	protein modification	142590
rs7214723	E375G	Possibly damaging	*CAMKK1*	calcium/calmodulin-dependent protein kinase kinase 1, α	nucleotide binding, kinase activity	
rs560191	D353E		*TP53BP1*	tumor protein p53 binding protein, 1	regulation of transcription	605230
rs3130618	R41L	Possibly damaging	*BAT4*	HLA-B associated transcript 4	nucleic acid binding	142610
rs2602141	K1136Q		*TP53BP1*	tumor protein p53 binding protein, 1	regulation of transcription	605230
rs6964587	M463I	Possibly damaging	*AKAP9*	A kinase (PRKA) anchor protein 9	receptor binding, signal transduction	604001
rs2229742	R448G	Possibly damaging	*NRIP1*	nuclear receptor interacting protein 1	receptor signaling, regulation of transcription	602490
rs2660744	Q488X	*Stop codon*	*PPAT*	phosphoribosyl pyrophosphate amidotransferase	transferase activity, nucleoside metabolism	172450
rs3206824	R335G		*DKK3*	dickkopf homolog 3	receptor signaling	605416
rs6183	P495T	Intolerant, Probably damaging	*GHR*	growth hormone receptor	endocytosis	600946
rs1129923	G131S	Intolerant	*DUSP23*	dual specificity phosphatase 23	protein tyrosine/serine/threonine phosphatase activity	
rs11571833	K3326X	*Stop codon*	*BRCA2*	breast cancer 2, early onset	DNA repair, regulation of transcription	600185
rs2242089	V105M	Intolerant	*PYCRL*	pyrroline-5-carboxylate reductase-like	oxidoreductase activity, electron transport	
rs1738023	N215D	Possibly damaging	*AKR7A3*	aldo-keto reductase family 7, member A3	aldehyde metabolism	608477
rs11569705	D22Y		*SULT1E1*	sulfotransferase family 1E, estrogen-preferring, member 1	steroid metabolism	600043
rs2295778	P41A	Possibly damaging	*HIF1AN*	hypoxia-inducible factor 1, α subunit inhibitor	regulation of transcription	606615
rs2306022	V433M	Intolerant	*ITGA11*	integrin, α 11	receptor signaling	604789
rs10115703	R19W		*CER1*	cerberus 1 homolog, cysteine knot superfamily	cell signaling	603777
rs2108978	R249H	Probably damaging	*AKAP10*	A kinase (PRKA) anchor protein 10	signal transduction	604694
rs2725362	L1074F		*WRN*	Werner syndrome	DNA metabolism	604611
rs3732401	P366S	Possibly damaging	*GTF2E1*	general transcription factor IIE, polypeptide 1, α 56 kDa	regulation of transcription	189962
rs4371716	V68M	Probably damaging	*CDH12*	cadherin 12, type 2 (N-cadherin 2)	cell adhesion	600562
rs8065506	K135N		*ZNF624*	zinc finger protein 624	regulation of transcription	
rs970547	G1894S		*COL12A1*	collagen, type XII, α 1	cell adhesion	120320
rs2032729	N220S		*ZNF24*	zinc finger protein 24 (KOX 17)	regulation of transcription	194534
rs2243639	T180A		*SFTPD*	surfactant, pulmonary-associated protein D	cell proliferation	178635
rs17632786	N700S	Intolerant, Possibly damaging	*THBS1*	thrombospondin 1	cell adhesion, cell motility	188060
rs10787428	E131G	Intolerant, Possibly damaging	*GPAM*	glycerol-3-phosphate acyltransferase, mitochondrial	lipid metabolism	602395
rs17184326	K522N		*POP1*	processing of precursor 1, ribonuclease subunit	tRNA catabolism	602486
rs11575194	R138W		*IGFBP5*	insulin-like growth factor binding protein 5	cell growth, signal transduction	146734
rs363504	L902S		*GRIK1*	glutamate receptor, ionotropic, kainate 1	cell signaling	138245
rs28360135	I134T		*XRCC4*	X-ray repair complementing defective repair in Chinese hamster cells 4	DNA repair, DNA recombination	194363
rs1051740	Y113H	Intolerant, Possibly damaging	*EPHX1*	epoxide hydrolase 1, microsomal	xenobiotic metabolism	132810
rs16900023	P786S		*MSH5*	mutS homolog 5 (*E. coli*)	DNA repair, DNA metabolism	603382
rs1043261	Q484X	*Stop codon*	*IL17RB*	interleukin 17 receptor B	cell growth	605458
rs2295000	S535L		*DATF1*	death associated transcription factor 1	apoptosis, regulation of transcription	604140
rs1019670	N150I	Intolerant	*MS4A6A*	membrane-spanning 4-domains, subfamily A, member 6A	signal transduction	606548

(continued)

Table 2. *Continued*

SNP	Substitution	Predicted Functionality[a]	Gene[b]	Gene Description	Gene Ontology[c]	OMIM[d]
rs7998427	E117G	Possibly damaging	*SETDB2*	SET domain, bifurcated 2	chromatin modification, methyltransferase activity	607865
rs9262138	D566G	Possibly damaging	*DHX16*	DEAH (Asp-Glu-Ala-His) box polypeptide 16	cell cycle control	603405
rs5745549	S914N		*MSH4*	mutS homolog 4 (*E. coli*)	DNA repair, meiotic recombination	602105
rs3107275	P421L	Intolerant	*PHACS*	1-aminocyclopropane-1-carboxylate synthase	transferase activity, amino acid metabolism	608405
rs8069344	L782H	Probably damaging	*GUCY2D*	guanylate cyclase 2D, membrane (retina specific)	cell signaling	600179
rs5388	V136I		*GH1*	growth hormone 1	signal transduction	139250
rs1800974	R651H		*ITGA7*	integrin, α 7	cell signaling	600536
rs1820128	C209R	Probably damaging	*ZNF600*	zinc finger protein 600	nucleic acid binding, zinc ion binding	
rs17246389	L6V	Intolerant	*SERPINI2*	serine proteinase inhibitor, clade I, member 2	cell motility	605587
rs17337252	M234T	Possibly damaging	*RB1CC1*	RB1-inducible coiled-coil 1	kinase activity	606837
rs3758938	K101T	Intolerant, Possibly damaging	*TBX10*	T-box 10	regulation of transcription	604648
rs1800076	R75Q	Intolerant, Possibly damaging	*CFTR*	CFT conductance regulator, ATP-binding (sub-family C, member 7)	ion transport	602421
rs2230674	P191A	Probably damaging	*ATF1*	activating transcription factor 1	regulation of transcription	123803
rs2274750	A1781T	Intolerant	*TNC*	tenascin C (hexabrachion)	cell adhesion	187380
rs933135	R257H		*PLCD1*	phospholipase C, δ 1	intracellular signaling, phospholipid metabolism	602142
rs5743611	R80T	Probably damaging	*TLR1*	toll-like receptor 1	regulation of TNF-α biosynthesis, macrophage activation	601194
rs1211554	D268Y		*HUS1B*	HUS1 checkpoint homolog b (*S. pombe*)	cell cycle control	
rs4647932	P464L	Possibly damaging	*FGFRL1*	fibroblast growth factor receptor-like 1	receptor activity	605830
rs17356233	H350Q	Possibly damaging	*CHD1L*	chromodomain helicase DNA binding protein 1-like	DNA repair	
rs3738888	R658C		*BARD1*	BRCA1 associated RING domain 1	apoptosis, DNA damage response	601593
rs12500797	E1606K	Intolerant	*PTPN13*	protein tyrosine phosphatase, non-receptor type 13	protein amino acid dephosphorylation	600267
rs4988492	L75F		*GHRH*	growth hormone releasing hormone	cell signaling, signal transduction	139190
rs2230339	R63Q	Intolerant	*GPR68*	G protein-coupled receptor 68	inflammatory response, signal transduction	601404
rs2229424	R1694H	Intolerant, Possibly damaging	*FASN*	fatty acid synthase	fatty acid biosynthesis	600212
rs4791641	P367L	Probably damaging	*PFAS*	phosphoribosyl-formylglycinamidine synthase (FGAR amidotransferase)	purine nucleotide biosynthesis	602133
rs11652709	Q122H	Probably damaging	*EPX*	eosinophil peroxidase	oxidative stress	131399
rs1801690	W335S	Probably damaging	*APOH*	apolipoprotein H (β-2-glycoprotein I)	cellular defense	138700

[a]Functional predictions based on SIFT (Intolerant) and PolyPhen (Probably damaging, Possibly damaging).
[b]NCBI Entrez Gene.
[c]Gene Ontology Database (http://www.geneontology.org).
[d]Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM).

death associated transcription factor 1 [*DATF1*]), or the DNA damage-response pathway (*BRCA2, MSH4, MSH5, XRCC4*), thereby having relevance to the pathobiology of lung cancer a priori.

There is evidence that several of the associated SNPs directly impact on the structure and function of the expressed protein, and are therefore likely to be directly responsible for the observed association. SNPs *BRCA2* K3326X and *THBS1* N700S are pre-eminent in this respect. The K3326X polymorphism in *BRCA2* results in loss of the terminal 91 amino acids of the expressed protein. The C-terminal region of *BRCA2* is involved in the nuclear colocalization of Fanconi anemia complementation group D2 (*FANCD2*) (Wang et al. 2004) and cells lacking the terminal 188 amino acids of *BRCA2* are hypersensitive to radiation (Morimatsu et al. 1998). SNP K3326X has been reported to play a role in *BRCA2*-related Fanconi's anemia (Howlett et al. 2002) and recently reported to increase the risk of pancreatic cancer (Martin et al. 2005). The N700S SNP of *THBS1,* encoding the anti-angiogenic protein thrombospondin, impacts on calcium binding vital for the normal function of THBS1, and has

been established to critically affect the structure and function of the expressed protein (Stenina et al. 2005).

For 37 SNPs correlated with lung cancer risk, evidence that they are deleterious is supported by predictions of functionality based on the PolyPhen and/or SIFT programs. Although in silico predictions about the functional consequences of amino acid changes are in part speculative, such algorithms have been demonstrated in benchmarking studies to successfully categorize 80% of amino acid substitutions (Xi et al. 2004). Two of these 37 putatively deleterious substitutions, A1718T (rs2274750) and P464L (rs4647932), were located in genes Tenascin C (*TNC*) and fibroblast growth-factor receptor-like 1 (*FGFRL1*), respectively. Both of these genes have been shown to be differentially expressed in the various lung cancer histologies (Garber et al. 2001) and form part of the extended GH-IGF pathway, with *FGFRL1* binding to fibroblast growth factor 2 (*FGF2*) and *TNC* interacting with epidermal growth factor receptor (*EGFR*) and *IGFBP5*.

Eleven of the 64 associated SNPs map to genes encoding pivotal components of the *GH-IGF1* pathways (Fig. 1). The absence of suitable nsSNPs in *AKT1, ARG2, FGF2, IGF1, PZDK1*, and *PRKCE* did not permit us to examine whether variants in these genes also contribute to lung cancer susceptibility. The prior probability of identifying a significant association with lung cancer risk for a series of 11 SNPs mapping to a single defined pathway of genes is intuitively small. The assertion that polymorphic variation and subsequent dysregulation in the *GH-IGF* axis could be associated with risk of lung cancer is not without precedent. *IGF1*, which is up-regulated by *GH*, regulates cellular proliferation and apoptosis and has been shown to increase tumor growth (Khandwala et al. 2000). Elevated levels of circulating *IGF1* have been shown to confer an increased risk of various tumors including breast (Toniolo et al. 2000), colorectal (Ma et al. 1999), lung (Yu et al. 1999), and prostate cancers (Chan et al. 1998). Furthermore, polymorphic variation in *IGFBP3* has been reported to increase risk of non-small cell lung cancer (NSCLC) (Moon et al. 2006). Recently, Bell et al. (2005) demonstrated that inherited susceptibility to lung cancer may be associated with acquisition of drug resistance mediated by *EGFR* T790M. While no studies have reported an association between *IGFBP5* variants and cancer to date, it is noteworthy that *IGFBP5* is required for regulation of cell-specific *IGF* responses during lung development (Schuller et al. 1995).

While it is desirable to validate our findings through analysis of additional large data sets, our study provides evidence that inherited predisposition to lung cancer is in part mediated through low-penetrance alleles and specifically identifies variants in genes comprising the *GH-IGF* pathway as susceptibility alleles.

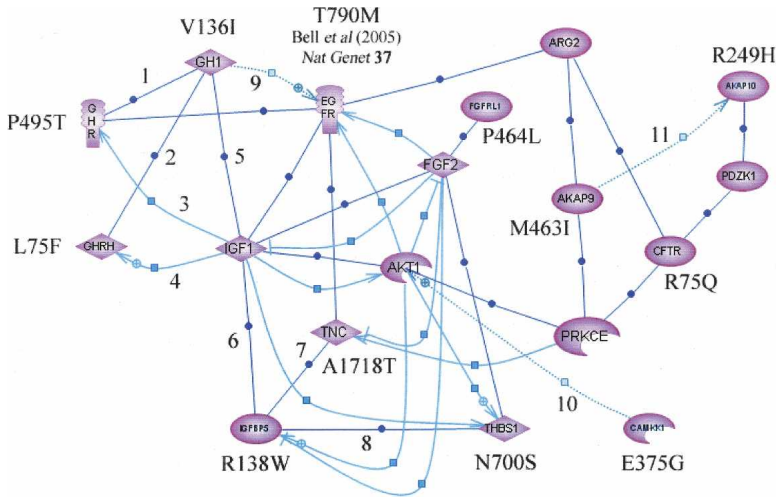

Figure 1. Inter-relationship between genes involved in the *GH-IGF* pathway containing SNPs associated with risk of lung cancer. Interactions were established using Pathway Assist software and are color-coded as follows: blue (expression), gray (regulation), and red (protein binding). Supporting publications are indicated with the corresponding NCBI Entrez PubMed ID in square brackets. (*1*) Binding [12,888,636]; (*2*) Binding [11,832,396]; (*3*) Expression [11,849,991]; (*4*) Expression [11,606,442]; (*5*) Binding [11,126,270]; (*6*) Binding [15,140,223]; (*7*) Binding [10,982,804]; (*8*) Binding [11,751,588]; (*9*) Regulation [14,517,795]; (*10*) Regulation [11,395,482]; (*11*) Regulation [15,047,863]. Validated nsSNPs with frequency data from Caucasian populations were not available in dbSNP Build 123 for genes *AKT1*, *ARG2*, *FGF2*, *IGF1*, *PZDK1*, and *PRKCE*. Bell et al. (2005) found association between lung cancer and SNP T790M.

Methods

Patients and control subjects

Patients with lung cancer were ascertained from the Genetic Lung Cancer Predisposition Study (GELCAPS) based in the United Kingdom (UK). Information on clinico-pathological characteristics and family history was collected using standardized questionnaires (Matakidou et al. 2005). In total, 1529 individuals with lung cancer were included in the study (506 males and 1023 females, median age at diagnosis 63 yr, range 26–92 yr). Case selection was prioritized firstly by family history and secondly, by early age-at-diagnosis. A total of 573 cases (38%) had a parent or sibling affected with lung cancer. Only 97 (6.3%) of 1529 cases were nonsmokers. Histology information was available for 1489 of the lung cancer cases—387 were small-cell cases and 1098 were non-small-cell cases (of which 483 were squamous and 343 adenocarcinomas).

A total of 2707 healthy individuals were recruited through either the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (1999–2004; http://intra-test.icr.ac.uk/tissueres/patient_blood.html), the National Study of Colorectal Cancer Genetics Trial (2004; http://www.ncrn.org.uk/portfolio/data.asp?ID=1269) or GELCAPS, all established within the UK. The control group contained 836 (31%) males and 1871 (69%) females, median age 59 yr (range 21–92 yr). None of the controls reported a personal history of cancer. All cases and controls were British Caucasians and there were no obvious demographic differences between groups in terms of place of residence within the UK. All study participants provided written informed consent. Ethical approval for the study was obtained from the London Multi-Center Research Ethics Committee (MREC/98/2/67) in accordance with the tenets of the Declaration of Helsinki. DNA was extracted from blood samples using conventional methodologies and quantified using PicoGreen (Invitrogen).

Selection of candidate genes and SNPs

We have previously established a publicly accessible PICS (Predicted Impact of

Coding SNPs) database (http://www.icr.ac.uk/cancgen/molgen/MolPopGen_PICS_database.htm) of potentially functional nsSNPs in genes with relevance to cancer biology (Rudd et al. 2005). Briefly, candidate cancer genes were identified by interrogating the Gene Ontology Consortium database (http://www.geneontology.org; Ashburner et al. 2000), Kyoto Encyclopedia of Genes and Genomes database (http://www.genome.jp/kegg; Kanehisa et al. 2004), Stratagene's Interaction Explorer Pathway Assist Program (http://www.iobion.com/news/hotnews.html?cmd=Retrieve&dopt=Abstract), National Center for Biotechnology Information (NCBI) Entrez Gene database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene; Maglott et al. 2005), and the CancerGene database (http://caroll.vjf.cnrs.fr/cancergene/HOME.html). A total of 9537 validated nsSNPs with minor allele frequency (MAF) data were identified within 21,506 LocusLink annotated genes in NCBI dbSNP Build 123 (http://www.ncbi.nlm.nih.gov/SNP/; Sherry et al. 2001). Filtering this list and linking it to 7080 candidate cancer genes yielded 3666 validated nsSNPs with MAF $\geq$ 0.01 in Caucasian populations. The functional impact of nsSNPs was predicted using the in silico computational tools PolyPhen (http://www.bork.embl-heidelberg.de/PolyPhen/; Ramensky et al. 2002) and SIFT (version 2.1; http://blocks.fhcrc.org/sift/SIFT.html; Ng and Henikoff 2001). Using the PICS database and published work on resequencing of DNA repair genes (Ford et al. 2000; Kuschel et al. 2002; Mohrenweiser et al. 2002; Fearnhead et al. 2004; Savas et al. 2004) we prioritized a set of 1476 nsSNPs for the current study. For those SNPs yet to be documented in the latest release of NCBI dbSNP (Build 125), we have submitted complete genotype information including MAF to NCBI and assigned the resultant dbSNP 'ss' designations accordingly. Annotated flanking sequence information for each SNP was derived from the University of California Santa Cruz (UCSC) Human Genome Browser (Assembly hg17; http://genome.ucsc.edu/cgi-bin/hgGateway).

SNP genotyping and data manipulation

Genotyping of samples was performed using customized Illumina Sentrix Bead Arrays according to the manufacturer's protocols. DNA samples with GenCall scores <0.25 at any locus were considered "no calls." A DNA sample was deemed to have failed if it generated genotypes at <95% of loci. A SNP was deemed to have failed if <95% of DNA samples generated a genotype at the locus. Conversion of genotype data into formats suitable for processing was performed using in-house Perl scripts (available upon request). Conventional statistical manipulations were undertaken in STATA (version 8; http://www.stata.com), S-Plus (version 7; http://www.insightful.com), or R (version 2.0.0; http://www.r-project.org).

Population stratification

Genotypic frequencies in control subjects for each SNP were tested for departure from Hardy-Weinberg equilibrium (HWE) using a χ^2 test or Fisher's exact test, where an expected cell count was less than five. SNPs that violate HWE in the control population can indicate selection bias or genotyping errors; these were removed from further analysis. To detect and control for possible population stratification, we used the genomic control approach (Devlin and Roeder 1999), using all SNPs to estimate the stratification parameter $\hat{\lambda}$ and its associated 95% confidence interval (CI).

Risk of lung cancer associated with nsSNPs

The most efficient test of association depends on the true mode of allelic inheritance. Since this is not known, we based our analyses on the difference between allelic frequencies in cases and controls using a χ^2 test with one degree of freedom, or Fisher's exact test if the expected numbers in individual cells were less than five. We denote this test statistic T_A with corresponding *P*-value P_A. We also investigated two further tests based on 2×2 tables combining the heterozygotes with either the common or rare homozygotes to derive the statistics T_R and T_D with corresponding *P*-values P_R and P_D, which are most powerful under recessive or dominant models, respectively. The risks associated with each SNP were estimated by allelic, dominant, and recessive odds ratios (ORs) using unconditional logistic regression. Associated 95% confidence intervals (CI) were calculated in each case. Where it was not possible to calculate ORs by asymptotic methods, an exact approach was implemented using LogXact software (http://www.cytel.com; Cytel Corporation).

To increase the power to detect associations, we further analyzed case and control genotypes adopting a set-association approach, combining the largest T_A statistics from individual tests into a single genome-wide statistic to model the joint effects of individual loci on lung cancer risk. Set-association analysis was conducted using the Sumstat program (Hoh et al. 2001), performing 50,000 iterations, and setting the maximum possible number of terms in the sum to be 100. The significance of this statistic was estimated through permutation, adjusting for the number of terms in the set being, a priori, unknown.

Multiple testing

Standard approaches to adjust for multiple testing such as the Bonferroni correction are known to be conservative due to their reliance on the assumption of independence between tests, which can lead to type I errors. To control error rate, we adopted an empirical Monte Carlo simulation approach (Churchill and Doerge 1994) based on 10,000 permutations, which takes into account the fact that tests may be correlated due to the presence of LD throughout the genome. At each iteration, case and control labels are permuted at random and maximum test statistics T_A^{max}, T_D^{max}, and T_R^{max} are determined. Significance levels of the observed statistics from the original data are then estimated by the proportion of permutation samples with T^{max} larger than that in the observed data. Although this approach adjusts for multiple testing for each of the three statistics separately, the consequent increase in false-positive rate is expected to be small due to the strong dependence between tests.

Assessment of linkage disequilibrium between SNPs

To identify SNPs in high LD, we calculated the pairwise LD measure D' between consecutive pairs of markers throughout the genome using the expectation–maximization algorithm to estimate two-locus haplotype frequencies. We computed D' for SNPs with MAF >0.1%, as the distribution of LD estimates for SNPs with smaller MAF was found to be unstable. For the purposes of this study, a pair of SNPs was defined as being in high LD if they had pairwise LD measure $D' > 0.5$. This information was used to investigate the relationship between haplotypes and disease status. Specifically, haplotypes were reconstructed using a Markov chain Monte Carlo method, and their frequencies in cases and controls compared by permutation testing using the program PHASE (Stephens et al. 2001; Stephens and Donnelly 2003).

Covariates and interactions

Information on a number of covariates was available for the cases, including family history of lung cancer, histology, age at diagnosis, smoking history, and asbestos exposure. Analyses were only undertaken for subgroups with a sample size >300. The test

statistics T_A, T_R, and T_D were computed for all subgroups, together with ORs and their associated 95% CIs. The set association approach was also implemented for each subgroup.

Under certain conditions, a two-stage process incorporating estimates of pairwise interaction between significant SNPs can yield greater power to detect association (Marchini et al. 2005). To investigate epistatic interactions, each pair of SNPs that showed significant allelic association at the 5% level were tested fitting a saturated logistic regression model, and the log likelihood ratio statistic for comparison with the main effects model computed. This was compared against a χ^2 distribution with 1 d.f. Statistics were then adjusted for multiple testing using a Bonferroni correction.

Acknowledgments

Funding for this work was undertaken with support from Cancer Research UK, the Arbib Foundation, HEAL, the National Cancer Research Network, the European Union Network of Excellence, and the Institute of Cancer Research. A.M. was the recipient of a clinical research fellowship from the Allan J. Lerner Fund. We gratefully acknowledge the participation of patient and control individuals. The authors are indebted to Richard Coleman, Christina Fleischmann, Olivia Fletcher, Nick Hearle, Nichola Johnson, Rosalind Mutch, Claire Palle, Julian Peto, Mobshra Qureshi, Elaine Ryder-Mills, Hayley Spendlove, and Remben Talaban for sample ascertainment. We thank Jurg Ott for access to a recompiled version of his Sumstat program.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25:** 25–29.

Bailey-Wilson, J.E., Amos, C.I., Pinney, S.M., Petersen, G.M., de Andrade, M., Wiest, J.S., Fain, P., Schwartz, A.G., You, M., Franklin, W., et al. 2004. A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am. J. Hum. Genet.* **75:** 460–474.

Bell, D.W., Gore, I., Okimoto, R.A., Godin-Heymann, N., Sordella, R., Mulloy, R., Sharma, S.V., Brannigan, B.W., Mohapatra, G., Settleman, J., et al. 2005. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat. Genet.* **37:** 1315–1316.

Chan, J.M., Stampfer, M.J., Giovannucci, E., Gann, P.H., Ma, J., Wilkinson, P., Hennekens, C.H., and Pollak, M. 1998. Plasma insulin-like growth factor-I and prostate cancer risk: A prospective study. *Science* **279:** 563–566.

Churchill, G.A. and Doerge, R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138:** 963–971.

Devlin, B. and Roeder, K. 1999. Genomic control for association studies. *Biometrics* **55:** 997–1004.

Doll, R. and Peto, R. 1981. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl. Cancer Inst.* **66:** 1191–1308.

Fearnhead, N.S., Wilding, J.L., Winney, B., Tonks, S., Bartlett, S., Bicknell, D.C., Tomlinson, I.P., Mortensen, N.J., and Bodmer, W.F. 2004. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc. Natl. Acad. Sci.* **101:** 15992–15997.

Ford, B.N., Ruttan, C.C., Kyle, V.L., Brackley, M.E., and Glickman, B.W. 2000. Identification of single nucleotide polymorphisms in human DNA repair genes. *Carcinogenesis* **21:** 1977–1981.

Garber, M.E., Troyanskaya, O.G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G.D., Perou, C.M., Whyte, R.I., et al. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci.* **98:** 13784–13789.

Hoh, J., Wille, A., and Ott, J. 2001. Trimming, weighting, and grouping SNPs in human case-control association studies. *Genome Res.* **11:** 2115–2119.

Howlett, N.G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., Persky, N., Grompe, M., Joenje, H., Pals, G., et al. 2002. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297:** 606–609.

Hwang, S.J., Cheng, L.S., Lozano, G., Amos, C.I., Gu, X., and Strong, L.C. 2003. Lung cancer risk in germline p53 mutation carriers: Association between an inherited cancer predisposition, cigarette smoking, and cancer risk. *Hum. Genet.* **113:** 238–243.

IARC (International Agency for Research on Cancer). 1986. *IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans: Tobacco smoking*. IARC, Lyon, France.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32:** D277–D280.

Khandwala, H.M., McCutcheon, I.E., Flyvbjerg, A., and Friend, K.E. 2000. The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. *Endocr. Rev.* **21:** 215–244.

Kuschel, B., Auranen, A., McBride, S., Novik, K.L., Antoniou, A., Lipscombe, J.M., Day, N.E., Easton, D.F., Ponder, B.A., Pharoah, P.D., et al. 2002. Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum. Mol. Genet.* **11:** 1399–1407.

Ma, J., Pollak, M.N., Giovannucci, E., Chan, J.M., Tao, Y., Hennekens, C.H., and Stampfer, M.J. 1999. Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. *J. Natl. Cancer Inst.* **91:** 620–625.

Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. 2005. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **33:** D54–D58.

Marchini, J., Donnelly, P., and Cardon, L.R. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.* **37:** 413–417.

Martin, S.T., Matsubayashi, H., Rogers, C.D., Philips, J., Couch, F.J., Brune, K., Yeo, C.J., Kern, S.E., Hruban, R.H., and Goggins, M. 2005. Increased prevalence of the BRCA2 polymorphic stop codon K3326X among individuals with familial pancreatic cancer. *Oncogene* **24:** 3652–3656.

Matakidou, A., Eisen, T., and Houlston, R.S. 2005. Systematic review of the relationship between family history and lung cancer risk. *Br. J. Cancer* **93:** 825–833.

Mohrenweiser, H.W., Xi, T., Vazquez-Matias, J., and Jones, I.M. 2002. Identification of 127 amino acid substitution variants in screening 37 DNA repair genes in humans. *Cancer Epidemiol. Biomarkers Prev.* **11:** 1054–1064.

Moon, J.W., Chang, Y.S., Ahn, C.W., Yoo, K.N., Shin, J.H., Kong, J.H., Kim, Y.S., Chang, J., Kim, S.K., Kim, H.J., et al. 2006. Promoter-202 A/C polymorphism of insulin-like growth factor binding protein-3 gene and non-small cell lung cancer risk. *Int. J. Cancer* **118:** 353–356.

Morimatsu, M., Donoho, G., and Hasty, P. 1998. Cells deleted for Brca2 COOH terminus exhibit hypersensitivity to γ-radiation and premature senescence. *Cancer Res.* **58:** 3441–3447.

Ng, P.C. and Henikoff, S. 2001. Predicting deleterious amino acid substitutions. *Genome Res.* **11:** 863–874.

Parkin, D.M., Bray, F., Ferlay, J., and Pisani, P. 2005. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55:** 74–108.

Ramensky, V., Bork, P., and Sunyaev, S. 2002. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res.* **30:** 3894–3900.

Rudd, M.F., Williams, R.D., Webb, E.L., Schmidt, S., Sellick, G.S., and Houlston, R.S. 2005. The predicted impact of coding single nucleotide polymorphisms database. *Cancer Epidemiol. Biomarkers Prev.* **14:** 2598–2604.

Sanders, B.M., Jay, M., Draper, G.J., and Roberts, E.M. 1989. Non-ocular cancer in relatives of retinoblastoma patients. *Br. J. Cancer* **60:** 358–365.

Savas, S., Kim, D.Y., Ahmad, M.F., Shariff, M., and Ozcelik, H. 2004. Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. *Cancer Epidemiol. Biomarkers Prev.* **13:** 801–807.

Schuller, A.G., van Neck, J.W., Beukenholdt, R.W., Zwarthoff, E.C., and Drop, S.L. 1995. IGF, type I IGF receptor and IGF-binding protein mRNA expression in the developing mouse lung. *J. Mol. Endocrinol.* **14:** 349–355.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, L., Phan, L., Smigielski, E.M., and Sirokin, K. 2001. dSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29:** 308–311.

Stenina, O.I., Ustinov, V., Krukovets, I., Marinic, T., Topol, E.J., and Plow, E.F. 2005. Polymorphisms A387P in thrombospondin-4 and N700S in thrombospondin-1 perturb calcium binding sites. *FASEB J.* **19:** 1893–1895.

Stephens, M. and Donnelly, P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73:** 1162–1169.

Stephens, M., Smith, N.J., and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68:** 978–989.

Takemiya, M., Shiraishi, S., Teramoto, T., and Miki, Y. 1987. Bloom's syndrome with porokeratosis of Mibelli and multiple cancers of the skin, lung and colon. *Clin. Genet.* **31:** 35–44.
Toniolo, P., Bruning, P.F., Akhmedkhanov, A., Bonfrer, J.M., Koenig, K.L., Lukanova, A., Shore, R.E., and Zeleniuch-Jacquotte, A. 2000. Serum insulin-like growth factor-I and breast cancer. *Int. J. Cancer* **88:** 828–832.
Wang, X., Andreassen, P.R., and D'Andrea, A.D. 2004. Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol. Cell. Biol.* **24:** 5850–5862.
Xi, T., Jones, I.M., and Mohrenweiser, H.W. 2004. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics* **83:** 970–979.
Yamanaka, A., Hirai, T., Ohtake, Y., and Kitagawa, M. 1997. Lung cancer associated with Werner's syndrome: A case report and review of the literature. *Jpn. J. Clin. Oncol.* **27:** 415–418.
Yu, H., Spitz, M.R., Mistry, J., Gu, J., Hong, W.K., and Wu, X. 1999. Plasma levels of insulin-like growth factor-I and lung cancer risk: A case-control analysis. *J. Natl. Cancer Inst.* **91:** 151–156.

Received January 6, 2006; accepted in revised form March 20, 2006.