Familial Aggregation of Common Sequence Variants on 15q24-25.1 in Lung Cancer

Pengyuan Liu, Haris G. Vikis, Daolong Wang, Yan Lu, Yian Wang,
Ann G. Schwartz, Susan M. Pinney, Ping Yang, Mariza de Andrade,
Gloria M. Petersen, Jonathan S. Wiest, Pamela R. Fain, Adi Gazdar,
Colette Gaba, Henry Rothschild, Diptasri Mandal, Teresa Coons,
Juwon Lee, Elena Kupert, Daniela Seminara, John Minna,
Joan E. Bailey-Wilson, Xifeng Wu, Margaret R. Spitz, Timothy Eisen,
Richard S. Houlston, Christopher I. Amos, Marshall W. Anderson, Ming You

Three recent genome-wide association studies identified associations between markers in the chromosomal region 15q24-25.1 and the risk of lung cancer. We conducted a genome-wide association analysis to investigate associations between single-nucleotide polymorphisms (SNPs) and the risk of lung cancer, in which we used blood DNA from 194 case patients with familial lung cancer and 219 cancer-free control subjects. We identified associations between common sequence variants at 15q24-25.1 (that spanned LOC123688 [a hypothetical gene], *PSMA4, CHRNA3, CHRNA5,* and *CHRNB4)* and lung cancer. The risk of lung cancer was more than five-fold higher among those subjects who had both a family history of lung cancer and two copies of high-risk alleles rs8034191 (odds ratio [OR] = 7.20, 95% confidence interval [CI] = 2.21 to 23.37) or rs1051730 (OR = 5.67, CI = 2.21 to 14.60, both of which were located in the 15q24-25.1 locus, than among control subjects. Thus, further research to elucidate causal variants in the 15q24-25.1 locus that are associated with lung cancer is warranted.

Lung cancer can occur sporadically in people with no known family history of lung cancer or it can occur in multiple members of the same family and be designated as familial lung cancer. Evidence of a genetic basis for susceptibility to lung cancer has been demonstrated through genome-wide association studies (1–3) and segregation analyses (4–9). We conducted a genome-wide association study among individuals with a familial history of lung cancer. These individuals are members of families with three or more first-degree relatives with lung cancer that were collected as part of the Genetic Epidemiology of Lung Cancer Consortium (GELCC).

For the genome-wide association study, we genotyped 194 case patients with familial lung cancer and 219 cancer-free control subjects by use of the Affymetrix 500K or Affymetrix Genome-Wide Human SNP Array 6.0 (Santa Clara, CA) (Supplementary Table 1, available online). To ensure genetic independence among subjects, one case patient with familial lung cancer was chosen from each high-risk lung cancer family. Noncancer control subjects were obtained from a combination of unaffected spouses from GELCC families (n = 36) and of unaffected individuals from the Coriell Institute for Medical Research (Camden, NJ) (n = 11) and the Fernald Medical Monitoring Program (Fernald, OH) (n = 172). These control subjects had no blood relationship with any selected case patients. To minimize possible effects of cigarette smoking and age, smokers with an older age were selected mainly as control subjects, except for spousal control subjects (Supplementary Table 1, available online). To maintain homogeneous population samples, only Caucasian subjects from the GELCC, Coriell Institute for Medical Research, and the Fernald Medical Monitoring Program were used for the association analysis. Basic characteristics of the GELCC subjects are presented in

Supplementary Table 1 (available online). The Texas and UK lung cancer cohorts have been described previously (1). Written informed consent and institutional review board approval were obtained for all the subjects involved in this study.

Single-nucleotide polymorphism (SNP) genotyping was done by the Vanderbilt University Microarray Shared Resource by following the Affymetrix protocol (www. affymetrix.com). We used the Affymetrix 500K chipset to genotype 76 control subjects and 73 case patients and the Affymetrix SNP 6.0 array to genotype the remaining 137 case patients and 164 control subjects. The samples used with the two platforms were randomly chosen from the GELCC collections. The Affymetrix 6.0 array includes more than 906 600 SNPs and covers 97.2% of SNPs presented in the Affymetrix 500K chipset of 500 568 SNPs. For the Affymetrix 500K SNP chipset, a confidence score of 0.33 was used for genotype determination, by using the Bayesian robust linear model with the Mahalanobis algorithm (10); an average genotyping call rate of 96.9% was obtained across all case and control samples. For the SNP Array 6.0, a block size of zero and a confidence threshold of 0.1 were used for genotype

CONTEXT AND CAVEATS

Prior knowledge
Genome-wide association studies have identified associations between markers in the chromosomal region 15q24-25.1 and the risk of lung cancer.

Study design
A genome-wide case–control association analysis was used to investigate relationships between single-nucleotide polymorphisms (SNPs) and the risk of familial lung cancer.

Contribution
Subjects with both a family history of lung cancer and two copies of either of two high-risk alleles in 15q24-25.1 had a higher risk of lung cancer than control subjects.

Implications
Additional research is required to identify which genetic variants in the 15q24-25.1 region are associated with a high risk of lung cancer.

Limitations
Associations of risk alleles with nicotine dependence were not directly tested because the data were not available. Smoking quantity was available; however, no association between smoking quantity and the high-risk alleles was found. The small sample size may have limited the ability to detect a smaller effect size for risk alleles among heterozygotes with familial lung cancer.

From the Editors

determination, by using the Birdseed algorithm (available at http://www.broad.mit.edu/mpg/birdsuite/birdseed.html). We excluded four samples with unexpected genetic relatedness that was detected by PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) and 33 samples with a genotype call rate of less than 86%. Thus, a total of 413 samples were available for our analysis (Supplementary Table 1, available online)—57 case patients and 56 control subjects genotyped with the Affymetrix 500K chipset, and 137 case patients and 163 control subjects genotyped with the Affymetrix SNP 6.0 array. When the frequency distribution of the overlapping SNPs in the two genotyping platforms was analyzed, no heterogeneity between the platforms was found (Supplementary Figure 1, available online). Among the 413

subjects, we detected no population stratification, as shown by linkage agglomerative clustering implemented in PLINK software. Hardy–Weinberg equilibrium for each SNP was examined with the software hweStrata, by use of an exact test as proposed previously (11) (stratum number $K = 1$). All statistical tests were two-sided. SNPs with an exact P value of .0001 or less from Hardy–Weinberg equilibrium tests and with a minor allele frequency of less than 1% among control subjects were excluded from association analyses. Thus, in the final association analysis of 413 samples, we retained 722 376 SNPs for the 300 samples that were genotyped by SNP Array 6.0 and 399 377 SNPs for the 113 samples that were genotyped by Affymetrix 500K. The statistical significance of the association between SNP allele and disease status was assessed primarily with Fisher exact tests (12) (Supplementary Figure 2, available online). The odds ratios (ORs) of lung cancer associated with each SNP and 95% confidence intervals (CIs) were estimated by allele and by genotype. We evaluated the cumulative distribution of P values from Fisher exact tests in our genome-wide association study. As shown in Supplementary Figure 3 (available online), the distribution of observed P values was similar to the expected uniform distribution [0, 1], indicating no inflation of test statistics from population structure or any other form of bias.

Our genome-wide association study identified several strong associations of SNPs on chromosomes 1, 3, 6, 9, 12, 15, and 20 with lung cancer. One of the most strongly associated clusters of SNPs was found in a 160-kb region at 15q24-25.1 (Figure 1, A and B, and Supplementary Figure 1, available online) that had strong linkage disequilibrium and contained multiple discrete haplotype blocks (ie, regions with no evidence of historical recombination), as defined by examining the HapMap data from Utah residents with ancestry from northern and western Europe (http://www.hapmap.org/) (Figure 1, C). These familial data confirm the recently described association between 15q25.1 and sporadic lung cancer (1–3). At least two common variants on 15q24-25.1, SNPs rs8034191 and rs1051730, were strongly associated with both familial ($P = 3.74 \times 10^{-4}$ for rs8034191 and $P = 2.90 \times 10^{-4}$ for rs1051730) and sporadic (from the

Texas study, $P = 9.71 \times 10^{-8}$ for rs8034191 and $P = 1.32 \times 10^{-7}$ for rs1051730; from the UK study, $P = 9.90 \times 10^{-9}$ for rs8034191 and $P = 2.84 \times 10^{-8}$ for rs1051730) lung cancer (Table 1). These SNPs are in high linkage disequilibrium ($r^2 = 0.87$) and are located within intronic regions of hypothetical gene LOC123688 (rs8034191) and of the gene *CHRNA3* (rs1051730). Risk of lung cancer among homozygous carriers was statistically significantly different from risk among noncarriers for the two SNPs—rs8034191 and rs1051730—in large sporadic lung cancer samples previously reported from Texas and from the United Kingdom (1).

It is worth noting that the effect size of the risk allele at 15q24-q25.1 observed in familial lung cancer (for carrying two copies of risk alleles of rs8034191 and rs1051730, respectively, OR = 3.84, 95% CI = 1.75 to 8.84, and OR = 3.43, 95% CI = 1.66 to 7.37) was larger than that observed in sporadic lung cancer (for carrying two copies of the same risk alleles, respectively, OR = 1.76, 95% CI = 1.42 to 2.18, and OR = 1.75, 95% CI = 1.42 to 2.17, in the Texas study; and OR = 2.09, 95% CI = 1.60 to 2.73, and OR = 2.05, 95% CI = 1.57 to 2.70, in the UK study) (Table 1). We therefore tested heterogeneity between familial and sporadic lung cancers by Woolf's test (13). Statistically significant heterogeneity was observed between familial and the Texas sporadic lung cancer samples ($P = .04$) but not between familial and UK sporadic lung cancer samples ($P = .12$).

To rule out confounding effects of smoking behavior on lung cancer risk, the association analysis was adjusted by sex, age, and pack-years of cigarette exposure, as continuous variables (Supplementary Table 2, available online). Association of the 15q24-25.1 locus with familial lung cancer remained statistically significant ($P = 1.03 \times 10^{-3}$ for rs8034191 and $P = 3.10 \times 10^{-4}$ for rs1051730) after this adjustment. The adjusted associations were even stronger among those in lung cancer families carrying two copies of risk alleles of rs8034191 (OR = 7.20, 95% CI = 2.21 to 23.37) or rs1051730 (OR = 5.67, 95% CI = 2.21 to 14.60). Thus, among lung cancer patients with a family history, alleles associated with a high risk of lung cancer appear to be located on the q arm of chromosome 15.

We next investigated whether the SNPs on 15q24-q25.1 that were strongly associated with lung cancer acted additively, recessively, or dominantly by use of logistic regression and Bayesian information criteria. Given any two estimated models, the model with the lower value of Bayesian information criteria is the one to be preferred. The strongest associations were found with the recessive model (Table 1 and Supplementary Table 3, available online). In the recessive model, an increased risk of lung cancer was associated with having two copies of the high-risk A allele of rs1051730, compared with having no copies (OR = 3.43, 95% CI = 1.66 to 7.37, P value = 2.90×10^{-4}). However, an additive model gave similar levels of statistical significance for many of the most statistically significant SNPs in this region. The homozygous risk genotype appeared to occur more frequently and to have larger effect size among familial lung cancer samples (OR = 3.84 for rs1051730) than among sporadic lung cancer samples (OR = 1.75 for rs1051730 from the Texas study and OR = 2.09 for rs1051730 in the UK study). The smaller sample size of the familial study was less powerful to detect the smaller risks among heterozygotes. This limitation and the higher frequency of homozygotes in the highly ascertained families with familial lung cancer may account for the difference in best fitting models at the 15q24 locus in familial vs sporadic data sets. A genome-wide association study on a larger sample of familial case patients would provide increased power. However, both familial and sporadic analyses support the presence of a risk allele at 15q24-25.1.

In addition to analyses of single SNPs, we performed a haplotype-based association analysis on the 15q24-q25.1 locus. Haplotypes at the 15q24-25.1 locus region were inferred by use of the fastPHASE computer program (14). To exhaustively exploit haplotype information, we then subject alleles (contiguous sets of markers) from sliding windows of all sizes to haplotype association tests (15). In the haplotype analysis, a common haplotype with a frequency larger than 5% was first identified and treated as one category and the other variants as another category. The Fisher exact test was then applied to determine statistical significance (12). Multiple haplotypes at the 15q24-25.1 locus were statistically significantly associated with familial lung cancer risk (Supplementary Table 4, available online). The most statistically significant association was with the haplotype A-T consisting of both rs7163730 and rs4461039 ($P = 1.03 \times 10^{-4}$). In the single-marker analyses, the two SNPs with the most statistically significant association with lung cancer were rs7163730 ($P = 8.71 \times 10^{-5}$) and rs4461039 ($P = 8.67 \times 10^{-5}$).

Using GELCC familial lung cancer samples and a genome-wide association study, we identified several common variants at 15q24-25.1 and confirmed the association between genetic variants on the q arm of chromosome 15 and sporadic lung cancer (1–3). When we adjusted for smoking and other covariates, a statistically significant association remained (Supplementary Table 2, available online). This observation is consistent with previous analyses (1–3), in which a direct role for these variants in lung cancer was postulated. In those analyses, the
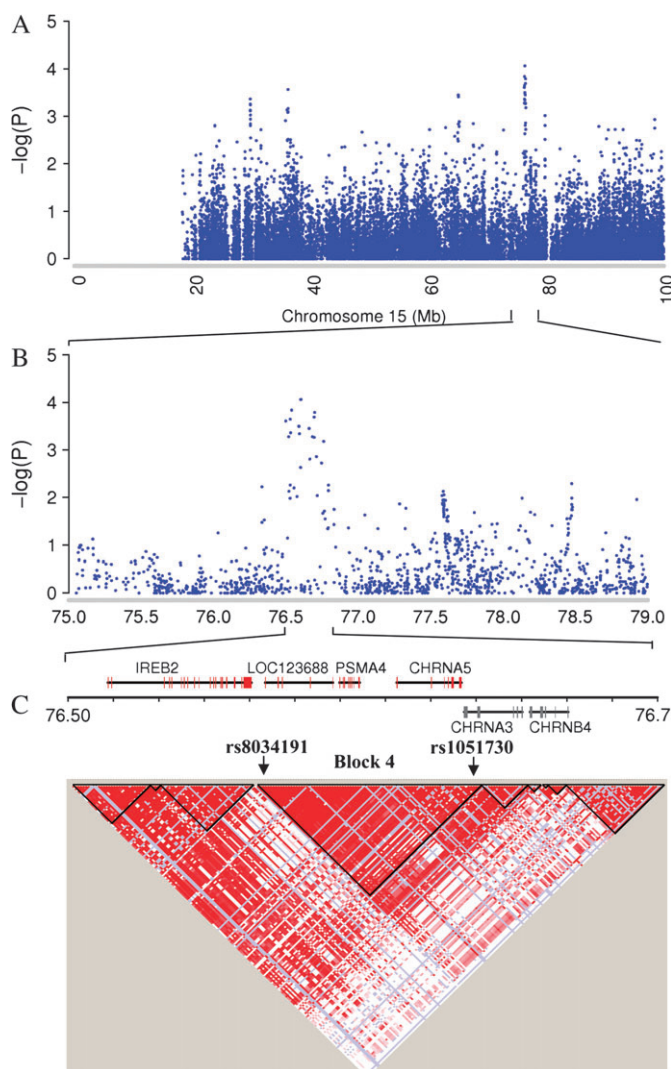

Figure 1. Association between chromosome 15q24-25.1 and lung cancer. Associations are expressed as –log(*P*); *P* values were from Fisher exact tests. All statistical tests were two-sided. **A**) Association analysis of 194 familial case patients and 219 control subjects in chromosome 15. **B**) Enhanced view of association analysis in the 75- to 79-Mb region of chromosome 15. Several common variants, including rs8034191 and rs1051730, cluster on 15q24-25.1 and show strong associations with familial lung cancer. **C**) Physical map and haplotype blocks (ie, regions with no evidence of historical recombination). Pairwise linkage disequilibrium, measured as *D'* (a measurement of the nonrandom association of alleles at two loci), was calculated with HapMap data from Utah residents with ancestry from northern and western Europe by the methods of Gabriel (21) as implemented with Haploview software (22). Shading represents the magnitude and statistical significance of the pairwise linkage disequilibrium, with a white-to-red gradient reflecting lower to higher linkage disequilibrium values. The single-nucleotide polymorphisms rs8034191 and rs1051730 were contained in the 84-kb block 4.

Table 1. Associations of the single-nucleotide polymorphisms rs8034191 and rs1051730 genotypes with lung cancer*

Samples†	Genotype	Frequency, %		P value‡	OR (95% CI)
		Control subjects	Case patients		
GELCC					
rs8034191	AA	50.00	36.64		
	AG	41.67	39.69	.362	1.30 (0.75 to 2.24)
	GG	8.33	23.66	3.74×10^{-4}	3.84 (1.75 to 8.84)
rs1051730§	GG	48.11	37.57		
	AG	44.81	43.35	.377	1.23 (0.78 to 1.96)
	AA	7.08	19.08	2.90×10^{-4}	3.43 (1.66 to 7.37)
Texas					
rs8034191	AA	44.10	37.01		
	AG	44.85	46.68	3.09×10^{-3}	1.24 (1.07 to 1.43)
	GG	11.05	16.32	9.71×10^{-8}	1.76 (1.42 to 2.18)
rs1051730	GG	44.31	36.82		
	AG	44.54	46.95	1.04×10^{-3}	1.27 (1.10 to 1.47)
	AA	11.15	16.23	1.32×10^{-7}	1.75 (1.42 to 2.17)
United Kingdom					
rs8034191	AA	46.67	36.59		
	AG	43.23	46.86	1.72×10^{-4}	1.38 (1.17 to 1.64)
	GG	10.10	16.55	9.90×10^{-9}	2.09 (1.60 to 2.73)
rs1051730	GG	46.55	37.54		
	AG	43.72	46.34	1.40×10^{-3}	1.31 (1.11 to 1.56)
	AA	9.73	16.12	2.84×10^{-8}	2.05 (1.57 to 2.70)

* GELCC = Genetic Epidemiology of Lung Cancer Consortium; OR = odds ratio; CI = confidence interval.

† The GELCC study had a total of 194 familial case patients and 219 control subjects. The Texas study had a total of 1865 sporadic case patients and 1769 control subjects. The UK study had a total of 2013 sporadic case patients and 3062 control subjects. The Texas and UK sporadic lung cancer samples are from the study by Amos et al. (1).

‡ Fisher exact tests were performed. All statistical tests were two-sided.

§ rs1051730 was genotyped by Human Genetics Division Genotyping Core.

number of cigarettes consumed per day, pack-years of exposure, and time to first cigarette were not associated with SNPs in the region (1,2). In contrast, Thorgeirsson et al. (3) reported that the SNP rs1051730 was statistically significantly associated with lung cancer and that each copy of the T allele was associated with an increase in smoking of one cigarette per day. In addition, Saccone et al. (16) identified the SNPs rs16969968 (missense in *CHRNA5*) and rs578776 (in the 3′-untranslated region of *CHRNA3*) as functional variants that were statistically significantly associated with nicotine dependence. When we genotyped rs16969968 and rs578776 in our familial lung cancer population, we found that both were statistically significantly associated with lung cancer (for rs16969968, $P = 2.29 \times 10^{-3}$; for rs578776, $P = 4.47 \times 10^{-4}$). Thus, polymorphisms in these genes may affect nicotine dependence and propensity to smoke and to develop lung cancer.

Our study had several limitations. First, the association of risk variants with nicotine dependence (as measured by the Fagerstrom test for nicotine dependence) was not directly tested because these data were unavailable. However, smoking quantity (a component of the Fagerstrom test) was available. We did not find an association between smoking quantity and the statistically significant variants associated with lung cancer. Second, the sample size in our familial genome-wide association analysis was small, and so we might not have been able to detect the smaller effect size of risk alleles located in 15q24-25.1 for heterozygotes in familial lung cancer.

The identity of variants in the 15q24-25.1 region that are most strongly associated with lung cancer remains unknown because many SNPs are in strong linkage disequilibrium with each other. In addition, whether the variants associate with a direct or indirect mode for lung cancer remains unresolved. From results of this and three other genome-wide association studies, the candidate genes include *IREB2*, LOC123688, *PSMA4*, *CHRNA5*, *CHRNA3*, and *CHRNB4* (1–3). *CHRNA5*, *CHRNA3*, and *CHRNB4* are strong candidates primarily because they encode subunits of the nicotinic acetylcholine receptors. These genes may partici-

pate in nicotinic addiction through reward pathways in the brain; however, evidence exists that nicotinic acetylcholine receptors are associated directly with lung carcinogenesis (17). Nicotinic acetylcholine receptors are expressed in lung cancers of both non–small cell and small cell subtypes (18), and treatment of lung cancer cell lines with nicotine can inhibit proapoptotic pathways initiated by opioids (19,20). Therefore, determination of a likely single candidate gene and further delineation of whether variants affect lung cancer directly or indirectly or both are warranted.

References

1. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008;40(5):616–622.
2. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature.* 2008;452(7187):633–637.
3. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008;452(7187):638–642.

4. Sellers TA, Bailey-Wilson JE, Elston RC, et al. Evidence for Mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst.* 1990;82(15):1272–1279.

5. Bailey-Wilson JE, Elston RC, Sellers TA, Rothschild H. Segregation analysis of lung cancer using class A regressive models. *Am J Hum Genet.* 1992;51:A145.

6. Sellers TA, Bailey-Wilson JE, Elston RC, Rothschild H. Evidence for Mendelian factors in early onset lung cancer. In: Brugge J, Carran T, Harlow E, McCormick F, eds. *Origins of Human Cancer: A Comprehensive Review.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1991:775–780.

7. Sellers TA, Potter JD, Folsom AR. Association of incident lung cancer with family history of female reproductive cancers: the Iowa Women's Health Study. *Genet Epidemiol.* 1991;8(3):199–208.

8. Sellers TA, Chen PL, Potter JD, Bailey-Wilson JE, Rothschild H, Elston RC. Segregation analysis of smoking-associated malignancies: evidence for Mendelian inheritance. *Am J Med Genet.* 1994;52(3):308–314.

9. Yang P, Schwartz AG, McAllister AE, Swanson GM, Aston CE. Lung cancer risk in families of nonsmoking probands: heterogeneity by age at diagnosis. *Genet Epidemiol.* 1999;17(4):253–273.

10. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics.* 2006;22(1):7–12.

11. Schaid DJ, Batzler AJ, Jenkins GD, Hildebrandt MA. Exact tests of Hardy-Weinberg equilibrium and homogeneity of disequilibrium across strata. *Am J Hum Genet.* 2006;79(6):1071–1080.

12. Ihaka R, Gentleman RR. a language for data analysis and graphics. *J Comput Graph Statist.* 1996;5(3):299–314.

13. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet.* 1955;19(4):251–253.

14. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 2006;78(4):629–644.

15. Lin S, Chakravarti A, Cutler DJ. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat Genet.* 2004;36(11):1181–1188.

16. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet.* 2007;16(1):36–49.

17. Egleton RD, Brown KC, Dasgupta P. Nicotinic acetylcholine receptors in cancer: multiple roles in proliferation and inhibition of apoptosis. *Trends Pharmacol Sci.* 2008;29(3):151–158.

18. Lam DC, Girard L, Ramirez R, et al. Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. *Cancer Res.* 2007;67(10):4638–4647.

19. Maneckjee R, Minna JD. Opioid and nicotine receptors affect growth regulation of human lung cancer cell lines. *Proc Natl Acad Sci USA.* 1990;87(9):3294–3298.

20. Maneckjee R, Minna JD. Opioids induce while nicotine suppresses apoptosis in human lung cancer cells. *Cell Growth Differ.* 1994;5(10):1033–1040.

21. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science.* 2002;296(5576):2225–2229.

22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–265.