Phase I Metabolic Genes and Risk of Lung Cancer: Multiple Polymorphisms and mRNA Expression

Melissa Rotunno1, Kai Yu1, Jay H. Lubin1, Dario Consonni3,4, Angela C. Pesatori3,4, Alisa M. Goldstein1, Lynn R. Goldin1, Sholom Wacholder1, Robert Welch2†, Laurie Burdette1,2, Stephen J. Chanock1,2, Pier Alberto Bertazzi3,4, Margaret A. Tucker1, Neil E. Caporaso1, Nilanjan Chatterjee1, Andrew W. Bergen1,5, Maria Teresa Landi1

1 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, 2 Core Genotyping Facility, Advanced Technology Program, Science Applications International Corporation-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland, United States of America, 3 Department of Occupational and Environmental Health, Clinica del Lavoro ‘L. Devoto’ University of Milan, Milan, Italy, 4 Department of Preventive Medicine, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Ospedale Maggiore Policlinico, Mangiagalli, Regina Elena Foundation, Milan, Italy, 5 Center for Health Sciences, SRI International, Menlo Park, California, United States of America

Abstract

Polymorphisms in genes coding for enzymes that activate tobacco lung carcinogens may generate inter-individual differences in lung cancer risk. Previous studies had limited sample sizes, poor exposure characterization, and a few single nucleotide polymorphisms (SNPs) tested in candidate genes. We analyzed 25 SNPs (some previously untested) in 2101 primary lung cancer cases and 2120 population controls from the Environment And Genetics in Lung cancer Etiology (EAGLE) study from six phase I metabolic genes, including cytochrome P450s, microsomal epoxide hydrolase, and myeloperoxidase. We evaluated the main genotype effects and genotype-smoking interactions in lung cancer risk overall and in the major histology subtypes. We tested the combined effect of multiple SNPs on lung cancer risk and on gene expression. Findings were prioritized based on significance thresholds and consistency across different analyses, and accounted for multiple testing and prior knowledge. Two haplotypes in EPHX1 were significantly associated with lung cancer risk in the overall population. In addition, CYP1B1 and CYP2A6 polymorphisms were inversely associated with adenocarcinoma and squamous cell carcinoma risk, respectively. Moreover, the association between CYP1A1 rs2606345 genotype and lung cancer was significantly modified by intensity of cigarette smoking, suggesting an underlying dose-response mechanism. Finally, increasing number of variants at CYP1A1/A2 genes revealed significant protection in never smokers and risk in ever smokers. Results were supported by differential gene expression in non-tumor lung tissue samples with down-regulation of CYP1A1 in never smokers and up-regulation in smokers from CYP1A1/A2 SNPs. The significant haplotype associations emphasize that the effect of multiple SNPs may be important despite null single SNP associations, and warrants consideration in genome-wide association studies (GWAS). Our findings emphasize the necessity of post-GWAS fine mapping and SNP functional assessment to further elucidate cancer risk associations.

Introduction

Lung cancer is the second most common malignancy and has the highest cancer mortality rate worldwide, with an estimated 161,840 individuals expected to succumb to the disease in 2008 in the US [1]. Tobacco smoking is the dominant causal factor for lung cancer; however, fewer than 20% of cigarette smokers develop the disease [2], suggesting that inherited genetic factors may also be important risk determinants. Genetic variation at tobacco carcinogen metabolizing enzymes may lead to inter-individual differences in the level of internal carcinogenic dose and to differential risk for individuals with similar exposures [3]. For this reason, genes that encode enzymes activating harmful chemicals are suitable candidates for lung cancer susceptibility studies and have been intensively studied [4]. Nevertheless, the available published data generally offer inconsistent results [5], due to population heterogeneity, low sample size, poor characterization of the exposure, and a few polymorphisms tested with low power to address the presence of their joint effects.

Here we addressed these issues in the analysis of candidate genes in phase I metabolism and lung cancer susceptibility, taking advantage of a large sample size and detailed epidemiological and clinical information of the Environment And Genetics in Lung cancer Etiology (EAGLE) study [6]. Furthermore, we integrated...
results on polymorphisms with data on expression from the same genes and the same subjects, for the first time in the context of a population study of phase I metabolic genes and lung cancer.

We explored the role of 25 single nucleotide polymorphisms (SNPs) covering important genes involved in the activation of carcinogens from cigarette smoking: cytochrome P450s (CYP1B1, CYP1A1, CYP1A2, and CYP2A6), microsomal epoxide hydrolase (EPHX1), and myeloperoxidase (MPO). We included also SNPs not previously analyzed, thus providing wide loci coverage in areas previously understudied.

Candidate genes

Many of the chemical carcinogens in tobacco smoke are members of the polycyclic aromatic hydrocarbon (PAH) family [7]. Cytochrome P450 enzymes activate PAHs [8] to epoxide intermediates, which are converted by epoxide hydrolase to the carcinogens diol-epoxides that interact with DNA or proteins to form adducts. In human lung for example, Benzo[a]pyrene (B[a]P) is a major carcinogenic constituent in tobacco smoke - it is first metabolically activated by cytochrome P450 1A1 (CYP1A1) and cytochrome P450 1B1 (CYP1B1) to form B[a]P-7,8-dihydropyrone-oxide, which is further hydrolyzed by microsomal epoxide hydrolase (EPHX1) to (F)-benzo[a]pyrene-trans-7,8-dihydriodiol. This compound is further metabolized by CYP1B1 to form benzo[a]pyrene-7,8-dihydriodil-9,10-epoxide [9], the most mutagenic and carcinogenic metabolite. CYP1A1 and CYP1B1 are over expressed in a wide range of human cancers, including breast, colon, lung, brain and testicular cancer [10,11]. Tobacco smoking can induce CYP1A1 and CYP1B1 proteins up to 10-fold higher levels, particularly in subjects (about 10% of the general population) that are more sensitive to enzyme induction [12]. Polymorphisms in CYP1A1 (chr15q24.1) are the most frequently studied in relation to lung cancer [13–17], but results are limited to only a few SNPs (rs1051740 and rs2294922, as reported by Kiyohara et al. in their review [38] and in more recent studies [39,40]. We included 8 SNPs from EPHX1 gene, 7 of which not previously studied in association with lung cancer. The human cytochrome P450 2A6 (CYP2A6) is responsible for the metabolism of different exogenous compounds including nitrosamines, aflatoxin B1, and other xenobiotic substrates [41]. In addition, CYP2A6 catalyzes nicotine C-oxidation to cotinine, and the subsequent hydroxylation of cotinine to 3-OH-cotinine [42]. Several genetic polymorphisms including point mutations and deletions have been reported and studied in association with lung cancer with conflicting results in populations from different ethnicities [43–45]. In particular the polymorphism CYP2A6 rs1801272 selected for this study, which causes an amino acid change from Leu to His, has been object of dispute: studies found a protective association with lung cancer and amount of cigarette smoke [46] which has not been consistently replicated.

Myeloperoxidase (MPO) gene, chr17q22 is a lysosomal enzyme present in high concentrations in human lung due to recruitment of neutrophils [47], and activates B[a]P [48] as well as aromatic amines [49] in tobacco smoke and generates carcinogen-free radicals [50]. A single base substitution, −463G>A, in the promoter region of MPO reduces transcription activity and DNA adduct levels in bronchoalveolar lavages of smokers [51]. These mechanisms have supported protective effects of the MPO −463A allele against lung cancer [52]. However, this possible inverse association with lung cancer risk has remained controversial [53]. Therefore, further study of the effects of this MPO polymorphism on lung cancer is warranted, and we included this SNP in our selection.

A precise characterization of the smoking exposure is essential to successfully identify molecular mechanisms involved in tobacco-related lung carcinogenesis. The EAGLE study provides detailed characterization of tobacco smoking including quantitative information on total exposure and daily intake of cigarette smoking. Using this information, we evaluated genotype-smoking interactions by likelihood ratio test, and compared the contributions of total exposure (pack-years) and intensity (cigarettes per day) of smoking using the linear-exponential model for smoking excess odds ratio (EOR) [54]. This model takes into account the correlation between the two smoking variables by describing the EOR per pack-year in terms of delivery rate of exposure. Our analyses also included stratified groups based on major lung cancer histology subtypes. Furthermore, we tested whether the overall lung cancer risk was determined by the combined action of multiple SNPs within the same gene, despite possible null effects in single SNP associations. We analyzed multiple SNPs jointly and performed gene haplotype analysis. The information on gene expression was limited to a subgroup of 44 subjects with adenocarcinoma, but can help clarify biological mechanisms behind the measured associations of lung cancer with polymorphisms in phase I metabolic genes. We prioritized our findings based on a low p-value threshold (p-value<0.01) and consistency across different analyses. In order to address concerns related to multiple testing and the priori knowledge considerations, we computed the False Positive Report Probability (FPRP) [55].
Results

Gene polymorphism and population characteristics

The 25 SNPs selected from phase I metabolic genes are presented in Table 1. The gene coverage is described in Supplemental Figure S1. All analyses were restricted to subjects with at least a 90% genotype call rate (i.e., 34 subjects were excluded). All 25 SNPs passed the test for Hardy-Weinberg equilibrium genotype proportions among the 2041 controls, with a p-value of 0.05 as the threshold.

Table 2 shows the frequency distributions and lung cancer association estimates for the main covariate, among the 4016 subjects included in the study. Age, sex and residential area were unrelated to case status, since frequency matching on these factors was in the design. As expected, all smoking related variables were associated with lung cancer, with increasing risks by increasing smoking exposures. Recent former smokers (up to 5 years) showed a higher risk for lung cancer compared to the current smokers. This is likely an artifact due to the fact that people typically quit smoking because of pre-clinical symptoms of lung cancer rather than a reflection of increasing risks in those who quit smoking [56]. In the analyses of genetic association we added the covariate “years since quit smoking” to the model, to adjust both for this reverse causation and for the attenuation of the risk over time.

SNP and lung cancer risk overall and by histology

Table 3 reports results with $p_{\text{trend}} \leq 0.05$ for the main effect associations between each SNP and lung cancer risk overall and by histology. The complete list of results is reported in Supplemental Table S1.

In adenocarcinoma cases only (test for heterogeneity by histology: $p_{\text{heterog}} = 0.066$), the minor allele of $CYP1B1$ rs10175368 was significantly protective for lung cancer (OR = 0.8, 95%CI = 0.69–0.93, $p_{\text{trend}} = 0.003$) and a similar protective effect was nominally significant (i.e. $p$-value $\leq 0.05$) for the $CYP1B1$ rs9341266 polymorphism. The cumulative number of variants in $CYP1B1$ rs9341266 and $CYP1B1$ rs10175368 also conferred a significant protection for lung cancer in adenocarcinoma cases only (OR = 0.83, 95%CI = 0.74–0.94, $p_{\text{trend}} = 0.002$; test for heterogeneity by histology: $p_{\text{heterog}} = 0.058$), in concordance with the two results from the single SNP analyses.

The $CYP2A6$ rs1801272 polymorphism was significantly associated with a decreased lung cancer risk in squamous cell carcinoma cases (OR = 0.47, 95%CI = 0.27–0.81, $p_{\text{trend}} = 0.007$; Table 2.

Table 1. List of studied genes, polymorphisms, and corresponding characteristics.

| Chromosome | Gene       | dbSNP (a) | SNP Region/Base Change (a) | AminoAcid Change (a) | Minor Allele (b) | MAF (b) |
|------------|------------|-----------|---------------------------|---------------------|------------------|---------|
| 1q42.12    | EPHX1      | rs2854455 | IVS1−1464T>C              |                     | C                | 0.251   |
|            |            | rs3766934 | IVS1−1409G>T              |                     | T                | 0.097   |
|            |            | rs2292566 | Ex3−8G>A                  | lys119lys          | A                | 0.138   |
|            |            | rs2260863 | IVS3+114C>G               |                     | G                | 0.326   |
|            |            | rs2234922 | Ex4+52A>G                 | his139Arg          | G                | 0.196   |
|            |            | rs34143170| Ex6+19C>T                 | his247His          | T                | 0.06    |
|            |            | rs2292568 | Ex6−80C>T                 | pro284Pro          | T                | 0.042   |
|            |            | rs1051741 | Ex8+31C>T                 | asn357Asn          | T                | 0.102   |
| 2p22.2     | CYP1B1     | rs163077  | *12295C>T                 |                     | T                | 0.217   |
|            |            | rs9341266 | Ex3−1249C>T (3’ UTR)      |                     | T                | 0.06    |
|            |            | rs162562  | Ex3+939A>C (3’ UTR)       |                     | C                | 0.157   |
|            |            | rs1800440 | Ex3+315A>G                | asn453ser          | G                | 0.201   |
|            |            | rs162557  | −2919C>T (upstream)       |                     | T                | 0.17    |
|            |            | rs162556  | −3922T>C (upstream)       |                     | C                | 0.446   |
|            |            | rs10175368| −5329G>A (upstream)       |                     | A                | 0.282   |
| 15q24.1    | CYP1A1     | rs2198843 | 11599 bp 3’ of STP G>C (intergenic) | C                | 0.17    |
|            |            | rs2606345 | IVS1+606T>G               |                     | G                | 0.358   |
|            |            | rs2470893 | −4010G>A (upstream)       |                     | A                | 0.204   |
|            |            | rs12441817| −10375A>G (intergenic)    |                     | G                | 0.079   |
|            |            | rs2472297 | −12441G>A (intergenic)    |                     | A                | 0.115   |
|            |            | rs2472299 | −17961C>T (intergenic)    |                     | T                | 0.321   |
| 15q24.1    | CYP1A2     | rs11072508| 14967 bp 3’ of STP T>C (intergenic) | C                | 0.388   |
|            |            | rs4886410 | *18214C>G (intergenic)    |                     | G                | 0.383   |
| 19q13.2    | CYP2A6     | rs1801272 | Ex3−15T>A                 | Leu160His          | A                | 0.041   |
| 17q22      | MPO        | rs2333227 | −642G>A (upstream) (aka −643 promoter) | A                | 0.255   |

(a) According to SNP500 database.
(b) Minor Allele and Minor Allele Frequency (MAF) are based on EAGLE controls.

doi:10.1371/journal.pone.0005652.t001

In adenocarcinoma cases only (test for heterogeneity by histology: $p_{\text{heterog}} = 0.066$), the minor allele of $CYP1B1$ rs10175368 was significantly protective for lung cancer (OR = 0.8, 95%CI = 0.69–0.93, $p_{\text{trend}} = 0.003$) and a similar protective effect was nominally significant (i.e. $p$-value $\leq 0.05$) for the $CYP1B1$ rs9341266 polymorphism. The cumulative number of variants in $CYP1B1$ rs9341266 and $CYP1B1$ rs10175368 also conferred a significant protection for lung cancer in adenocarcinoma cases only (OR = 0.83, 95%CI = 0.74–0.94, $p_{\text{trend}} = 0.002$; test for heterogeneity by histology: $p_{\text{heterog}} = 0.058$), in concordance with the two results from the single SNP analyses.

The $CYP2A6$ rs1801272 polymorphism was significantly associated with a decreased lung cancer risk in squamous cell carcinoma cases (OR = 0.47, 95%CI = 0.27–0.81, $p_{\text{trend}} = 0.007$;
Table 2. Characteristics of lung cancer cases and controls from the EAGLE population with genotype call rate ≥ 90%, and their association with lung cancer status.

| Characteristic | Sub-category | Cases | Controls | Association with case:control status |
|----------------|--------------|-------|----------|--------------------------------------|
|                |              | n     | %        | n         | %         | OR [95%CI] # |
| Sex            | Males        | 1563  | 79.1     | 1560      | 76.4      | 1.0         |
|                | Females      | 412   | 20.9     | 481       | 23.6      | 0.88 [0.76–1.03] |
| Age            | 35–39        | 11    | 0.6      | 15        | 0.7       | 1.0         |
|                | 40–44        | 17    | 0.9      | 26        | 1.3       | 0.88 [0.33–2.39] |
|                | 45–49        | 50    | 2.5      | 67        | 3.3       | 1.03 [0.44–2.46] |
|                | 50–54        | 124   | 6.3      | 121       | 5.9       | 1.39 [0.61–3.18] |
|                | 55–59        | 222   | 11.2     | 289       | 14.2      | 1.03 [0.46–2.31] |
|                | 60–64        | 337   | 17.1     | 356       | 17.4      | 1.25 [0.56–2.77] |
|                | 65–69        | 445   | 22.5     | 472       | 23.1      | 1.24 [0.56–2.75] |
|                | 70–74        | 442   | 22.4     | 412       | 20.2      | 1.42 [0.64–3.15] |
|                | 75–79        | 327   | 16.6     | 283       | 13.9      | 1.56 [0.70–3.48] |
| Area           | Brescia      | 261   | 13.2     | 240       | 11.8      | 1.0         |
|                | Milan        | 1302  | 65.9     | 1389      | 68.1      | 0.85 [0.71–1.04] |
|                | Monza        | 133   | 6.7      | 111       | 5.4       | 1.10 [0.81–1.50] |
|                | Pavia        | 126   | 6.4      | 122       | 6.0       | 0.96 [0.71–1.30] |
|                | Varese       | 153   | 7.7      | 179       | 8.8       | 0.78 [0.59–1.04] |
| Smoking status | Never        | 140   | 7.1      | 658       | 32.2      | 1.0         |
|                | Former, >2 years | 655   | 33.2     | 848       | 41.5      | 3.98 [3.18–4.98] |
|                | Former, 0.5 to 2 years | 188   | 9.5      | 30        | 1.6       | 34.37 [22.22–53.16] |
|                | Current      | 980   | 49.6     | 501       | 24.5      | 11.37 [9.06–14.28] |
|                | Missing      | 12    | 0.6      | 4         | 0.2       |
| Cigarettes per day | Never      | 140   | 7.1      | 658       | 32.2      | 1.0         |
|                | <12          | 233   | 11.8     | 519       | 25.4      | 2.54 [1.98–3.26] |
|                | 12–20        | 358   | 18.1     | 343       | 16.8      | 7.00 [5.41–9.05] |
|                | 20–25        | 571   | 28.9     | 290       | 14.2      | 14.77 [11.37–19.18] |
|                | >25          | 568   | 28.8     | 226       | 11.1      | 19.63 [14.98–25.72] |
|                | Missing      | 105   | 5.3      | 5         | 0.2       |
| Total pack-years | Never      | 140   | 7.1      | 658       | 32.2      | 1.0         |
|                | <19.5        | 187   | 9.5      | 578       | 28.3      | 1.88 [1.45–2.43] |
|                | 19.5–36      | 381   | 19.3     | 365       | 17.9      | 7.25 [5.61–9.37] |
|                | 36–52.5      | 539   | 27.3     | 275       | 13.5      | 15.02 [11.54–19.56] |
|                | >52.5        | 623   | 31.5     | 160       | 7.8       | 30.79 [23.19–40.86] |
|                | Missing      | 105   | 5.3      | 5         | 0.2       |
| Years since quit | Current    | 980   | 49.6     | 501       | 24.5      | 1.0         |
|                | <5           | 300   | 15.2     | 96        | 4.7       | 1.49 [1.15–1.93] |
|                | 5–15         | 249   | 12.6     | 179       | 8.8       | 0.63 [0.50–0.78] |
|                | 15–24        | 180   | 9.1      | 260       | 12.7      | 0.31 [0.25–0.39] |
|                | >24          | 114   | 5.8      | 343       | 16.8      | 0.14 [0.11–0.18] |
|                | Never        | 140   | 7.1      | 658       | 32.2      | 0.09 [0.07–0.11] |
|                | Missing      | 12    | 0.6      | 4         | 0.2       |
test for heterogeneity by histology: $p_{\text{interact}} = 0.045$. The protective effect was nominally significant in the overall population. Interestingly, the same SNP was significantly associated with a decrease of cigarette smoking intensity in controls (OR = 0.86, 95%CI = 0.78–0.94, $p_{\text{trend}} = 0.0007$).

### Genotype-smoking interaction

We repeated the analyses within subgroups defined by smoking status (never and ever smokers) in all cases and controls and, separately, in adenocarcinoma cases only and all controls (see Table 4 for the single SNP analysis, and Supplemental Table S2 for the joint SNP analysis). The other histology groups included too few never smokers to perform a meaningful analysis.

Three SNPs in the chr15q24.1 region ($\text{CYP1A1/A2}$) showed a protective effect for lung cancer among never smokers but a tendency towards increased risk of lung cancer in ever smokers, with a significant genotype-smoking interaction for $\text{CYP1A1}$ rs2606345 ($p_{\text{interact}} = 0.005$) and a nominally significant genotype-smoking interaction for the two SNPs in $\text{CYP1A2}$.

We further explored the significant genotype-smoking interaction in $\text{CYP1A1}$ rs2606345 by means of the linear-exponential model for smoking excess odds ratio [54], and evaluated whether the variation in smoking risk by genotype resulted from the interaction with smoking intensity or with total pack-years and whether this interaction was present among other categories of smokers such as current or former smokers. Results are shown in Figure 1. The EOR per pack-years in current smokers compared to never smokers (Figure 1A and Figure 1B) increased for increasing number of cigarettes per day, reaching a plateau for subjects carrying the $\text{CYP1A1}$ rs2606345 homozygote major allele (Figure 1A), and in contrast, increasing exponentially for subjects carrying the $\text{CYP1A1}$ rs2606345 heterozygote or homozygote minor allele (Figure 1B). The same analysis of EOR/pack-years in former smokers versus never smokers (Figure 1C and Figure 1D) similarly showed that the EOR increase for cigarettes per day was lower in homozygote major allele carriers (Figure 1C) than for heterozygote or homozygote minor alleles carriers (Figure 1D), but here EOR/pack-years reached a plateau among both groups of subjects. Panel E in Figure 1 reports the estimated deviances and p-values for the genotype-smoking interaction among current and former smokers for the model including both interaction terms between the genotype and pack-years and between the genotype and cigarettes per day, and for intermediate models including either the interaction term between genotype and pack-years, or the interaction term between genotype and cigarettes per day. The overall genotype-smoking interaction was stronger among current smokers ($p_{\text{interact}} = 0.009$) than among former smokers ($p_{\text{interact}} = 0.124$). Among current smokers, the removal of pack-years from the model did not degrade fit relative to the full model ($p = 0.209$), whereas the removal of cigarettes per day did degrade fit ($p = 0.022$), suggesting that the genotype interaction effects resulted from cigarettes per day and not pack-years.

In the joint analysis of multiple SNPs stratified by smoking (Supplemental Table S2), the cumulative number of variants of all 8 SNPs from $\text{CYP1A1}$ and $\text{CYP1A2}$ in the chr15q24.1 region conferred a significant overall risk for lung cancer in ever smokers (OR = 1.03, 95%CI = 1.00–1.07, $p_{\text{trend}} = 0.040$) and a borderline protective effect in never smokers (OR = 0.91, 95%CI = 0.84–0.99, $p_{\text{trend}} = 0.055$). The smoking-genotype interaction was highly significant ($p_{\text{interact}} = 0.006$).

In addition, the minor allele of $\text{CYP2A6}$ rs1801272 showed a significant protective effect in ever smokers, increased risk in never smokers, and a nominally significant genotype-smoking interaction.

### Linkage disequilibrium and haplotype analysis

For genes represented by two or more SNPs, we computed linkage disequilibrium (LD) among controls and haplotype association with lung cancer. The complete results are reported in the Supplemental Text S1 and Figure S2.

Interestingly, the haplotype analysis for the 8 SNPs in $\text{EPHX1}$ (which were in low LD: $r^2 \leq 0.1$ for most SNPs pairs, $r^2 = 0.43$ for $\text{EPHX1}$ rs2234922 and $\text{EPHX1}$ rs1051741) revealed two haplotypes significantly associated with lung cancer in the overall population: carriers of $\text{TGGCAGCT}$ haplotype had higher risk than non-carriers (freq = 0.01, p-value = 0.010) and carriers of $\text{GGGC-GCCT}$ haplotype had a lower risk than non-carriers (freq = 0.01, p-value = 0.015). In addition, we found similar results in the analysis restricted to adenocarcinoma cases only: $\text{TGGCAGCT}$ (p-value = 0.008) and $\text{GGGC-GCCT}$ (p-value = 0.025). Since the 8 SNPs were
in low LD, we also performed a three marker moving window haplotype analysis and found no significant associations between lung cancer and haplotype combinations of three SNPs (see Supplemental Table S3). However, we identified a borderline significant protective association (freq = 0.03, p-value = 0.059) with a three-locus haplotype with a C, G, and T in locus 1, 2 and 8, respectively, which was also contained in the 8 SNP haplotype.

For the 8 SNPs in the chr15q24.1 region, we found two regions of LD, one of modest strength surrounding CYP1A1, and a second region around CYP1A2 (see Supplemental Figure S2), concordant with the results from HapMap. Haplotype analyses were computed separately for these two LD regions; the GTAAA haplotype (freq = 0.07) and the CGGGG haplotype (freq = 0.03) were nominally significantly associated with lung cancer risk in never and ever smokers respectively.

Association between genotype and gene expression

The complete results for the correlation between genotype and gene expression data are reported in Supplemental Table S4. We found that the 8 polymorphisms in the 15q24 chromosomal region had a significant down-regulating effect on mRNA expression for CYP1A1 gene among the 14 never smokers ($d = 2.151$, p-value = 0.007) and showed a trend for up-regulation among the 15 current smokers ($d = 4.95$, p-value = 0.078). The 7 polymorphisms in CYP1B1 were significantly associated with an increase of mRNA expression in CYP1B1 among the 15 current smokers ($d = 8.99$, p-value = 0.004), and not among the 44 subjects overall. For the 8 SNPs in EPHX1 gene, we observed an overall trend for decreasing expression ($d = -2.56$, p-value = 0.049).

### Table 3. Polymorphisms associated with risk of lung cancer overall and by histology with a significant trend (in bold) or nominally significant trend (in *italics*).

| SNP        | Genotype | Controls | Cases | OR (a) | 95%CI−   | 95%CI+   | P-value Trend# |
|------------|----------|----------|-------|--------|----------|----------|---------------|
| **All Histologies** |          |          |       |        |          |          |               |
| CYP2A6/rs1801272 | T/T      | 1855     | 1756  | 1      |          |          |               |
|             | T/A      | 160      | 101   | 0.74   | 0.55     | 1.00     |               |
|             | A/A      | 4        | 2     | 0.26   | 0.04     | 1.94     |               |
|             | T+A/A/A  | 164      | 103   | 0.73   | 0.54     | 0.98     |               |
|             | Trend    |          |       | 0.72   | 0.54     | 0.96     | 0.026         |
| EPHX1/rs2292568 | C/C      | 1852     | 680   | 1      |          |          |               |
|             | C/T      | 156      | 86    | 1.48   | 1.09     | 2.01     |               |
|             | T/T      | 7        | 1     | 0.41   | 0.04     | 4.43     |               |
|             | C/T+C/T  | 163      | 87    | 1.44   | 1.06     | 1.96     |               |
|             | Trend    |          |       | 1.38   | 1.03     | 1.85     | 0.032         |
| CYP1B1/rs9341266 | C/C      | 1798     | 701   | 1      |          |          |               |
|             | C/T      | 222      | 72    | 0.8    | 0.59     | 1.09     |               |
|             | T/T      | 12       | 1     | 0.14   | 0.01     | 1.24     |               |
|             | C/T+C/T  | 234      | 73    | 0.76   | 0.56     | 1.04     |               |
|             | Trend    |          |       | 0.74   | 0.55     | 0.99     | 0.046         |
| CYP1B1/rs162556 | T/T      | 621      | 205   | 1      |          |          |               |
|             | T/C      | 1002     | 391   | 1.15   | 0.92     | 1.42     |               |
|             | C/C      | 400      | 172   | 1.34   | 1.03     | 1.74     |               |
|             | T/C+C/C  | 1402     | 563   | 1.2    | 0.98     | 1.47     |               |
|             | Trend    |          |       | 1.16   | 1.01     | 1.32     | 0.031         |
| CYP1B1/rs10175368 | G/G      | 1056     | 430   | 1      |          |          |               |
|             | G/A      | 790      | 297   | 0.87   | 0.71     | 1.05     |               |
|             | A/A      | 176      | 45    | 0.55   | 0.38     | 0.81     |               |
|             | G/A+A/A  | 966      | 342   | 0.81   | 0.67     | 0.97     |               |
|             | Trend    |          |       | 0.8    | 0.69     | 0.93     | 0.003         |
| **Squamous Cell Carcinoma** | T/T      | 1855     | 463   | 1      |          |          |               |
|             | T/A      | 160      | 18    | 0.48   | 0.27     | 0.86     |               |
|             | A/A      | 4        | 0     | -      | -        | -        |               |
|             | T+A/A/A  | 164      | 18    | 0.47   | 0.27     | 0.83     |               |
|             | Trend    |          |       | 0.47   | 0.27     | 0.81     | 0.007         |

(a) ORs were adjusted for age, sex, area, cigarette per day, total pack-years, years since quit smoking.

#Two-sided Wald test.

doi:10.1371/journal.pone.0005652.t003
Table 4. Associations between SNPs and lung cancer risk by never/ever smoking status, for significant (in bold) or nominally significant (in italics) smoking-genotype interactions.

| SNP                | Genotype | Never Smokers |               |               |               |               | Ever Smokers |               |               |               |               |               |               |               |               |               |               |               |               |               |
|--------------------|----------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                    |          | Controls      | Cases         | OR (a)        | 95% CI −      | 95% CI +      | Trend         | Controls      | Cases         | OR (b)        | 95% CI −      | 95% CI +      | P-val Trend, # | LH Ratio      | P-value       |               |               |               |               |               |               |
| All Histologies    |          |               |               |               |               |               |               |               |               |               |               |               |               |               |               |               |               |               |               |               |
| CYP1A1/rs2606345   | T/T      | 262           | 73            | 1             |               |               |               | 579           | 724           | 1             |               |               |               |               |               |               |               |               |               |               |
|                    | T/G      | 300           | 54            | 0.68          | 0.46          | 1.01          |               | 622           | 769           | 0.99          | 0.84          | 1.18          |               |               |               |               |               |               |               |
|                    | G/G      | 95            | 13            | 0.48          | 0.25          | 0.92          |               | 170           | 233           | 1.18          | 0.92          | 1.53          |               |               |               |               |               |               |               |
|                    | T/G+G/G  | 395           | 67            | 0.63          | 0.43          | 0.91          |               | 792           | 1002          | 1.03          | 0.88          | 1.21          |               |               |               |               |               |               |               |
| Trend              |          |               |               | 0.69          | 0.52          | 0.91          | **0.009**     |               |               | 1.06          | 0.94          | 1.19          | 0.334          | **0.005**     |               |               |               |               |               |               |
| CYP1A2/rs11072508  | T/T      | 247           | 61            | 1             |               |               |               | 515           | 627           | 1             |               |               |               |               |               |               |               |               |               |               |
|                    | T/G      | 299           | 61            | 0.86          | 0.58          | 1.29          |               | 664           | 803           | 1.04          | 0.87          | 1.24          |               |               |               |               |               |               |               |
|                    | G/G      | 110           | 17            | 0.65          | 0.36          | 1.18          |               | 196           | 292           | 1.31          | 1.02          | 1.67          |               |               |               |               |               |               |               |
|                    | T/G+G/G  | 409           | 78            | 0.81          | 0.55          | 1.18          |               | 860           | 1095          | 1.1           | 0.93          | 1.3           |               |               |               |               |               |               |               |
| Trend              |          |               |               | 0.82          | 0.63          | 1.08          | **0.157**     |               |               | 1.12          | 1             | 1.26          | 0.055          | 0.038         |               |               |               |               |               |               |
| CYP1A2/rs4886410   | C/C      | 249           | 61            | 1             |               |               |               | 521           | 640           | 1             |               |               |               |               |               |               |               |               |               |               |
|                    | C/G      | 300           | 62            | 0.85          | 0.57          | 1.27          |               | 665           | 805           | 1.04          | 0.87          | 1.24          |               |               |               |               |               |               |               |
|                    | G/G      | 108           | 17            | 0.67          | 0.37          | 1.22          |               | 187           | 281           | 1.3           | 1.01          | 1.67          |               |               |               |               |               |               |               |
|                    | C/G+G/G  | 408           | 79            | 0.8           | 0.55          | 1.17          |               | 852           | 1086          | 1.09          | 0.93          | 1.29          |               |               |               |               |               |               |               |
| Trend              |          |               |               | 0.83          | 0.63          | 1.09          | **0.175**     |               |               | 1.12          | 0.99          | 1.25          | 0.069          | 0.047         |               |               |               |               |               |               |
| CYP2A6/rs1801272   | T/T      | 601           | 124           | 1             |               |               |               | 1254          | 1632          | 1             |               |               |               |               |               |               |               |               |               |               |
|                    | T/A      | 47            | 16            | 1.51          | 0.81          | 2.79          | 91            | 113           | 85            | 0.62          | 0.44          | 0.87          |               |               |               |               |               |               |               |
|                    | A/A      | 0             | 0             | -             | -             | -             | 4             | 2             | 0.26          | 0.03          | 1.92          |               |               |               |               |               |               |               |
|                    | T/A+A/A  | 47            | 16            | 1.51          | 0.81          | 2.79          | 117           | 87            | 0.61          | 0.44          | 0.84          |               |               |               |               |               |               |               |
| Trend              |          |               |               | 1.51          | 0.81          | 2.79          | **0.191**     |               |               | 0.61          | 0.44          | 0.84          | **0.002**      | **0.026**    |               |               |               |               |               |               |
| Adenocarcinoma     | CYP1A1/rs2606345 | T/T     | 262           | 51            | 1             |               |               | 579           | 281           | 1             |               |               |               |               |               |               |               |               |               |               |
|                    | T/G      | 300           | 38            | 0.68          | 0.43          | 1.08          |               | 622           | 286           | 0.94          | 0.76          | 1.17          |               |               |               |               |               |               |               |
|                    | G/G      | 95            | 11            | 0.59          | 0.29          | 1.2           |               | 170           | 107           | 1.39          | 1.02          | 1.89          |               |               |               |               |               |               |               |
|                    | T/G+G/G  | 395           | 49            | 0.66          | 0.43          | 1.01          |               | 792           | 393           | 1.03          | 0.84          | 1.27          |               |               |               |               |               |               |               |
| Trend              |          |               |               | 0.74          | 0.54          | 1.02          | **0.066**     |               |               | 1.11          | 0.96          | 1.29          | 0.154          | **0.022**    |               |               |               |               |               |               |

(a) ORs were adjusted for age, sex, area.
(b) ORs were adjusted for age, sex, area, cigarette per day, total pack-years, years since quit smoking.
(1) Two-sided Wald test.
doi:10.1371/journal.pone.0005652.t004

Phase I Metabolic Genes
Multiple testing

FPRP calculations (Table 5) were performed for nominally significant or significant single SNP analysis results. The table shows that all prior probabilities of ≥0.10 had low FPRP values (<0.5).

Discussion

In this large population-based case-control study of lung cancer, we have observed that *EPHX1, CYP1A1, CYP1B1* and *CYP2A6* genes may play a role in lung cancer susceptibility.
A group of SNPs from two LD regions in the chr15q24.1 region (CYP1A1 and CYP1A2) showed a protective effect on lung cancer risk among never smokers and a suggestive risk of lung cancer in ever smokers with a significant genotype-smoking interaction for CYP1A1 rs2606345 and a nominally significant interaction for the two SNPs in CYP1A2. This result was confirmed by the multiple SNP analysis stratified by smoking. The cumulative number of variants from CYP1A1 and CYP1A2 was in fact associated with a significant risk for lung cancer in ever smokers and a protective effect in never smokers, with a highly significant smoking-genotype interaction. Interestingly, Wang et al. [58] recently reported an analogous inverse association between CYP1A1 rs2606345 and levels of DNA adducts: the variant allele was associated with high level of DNA adducts among women with high PAH exposure and with low level of DNA adducts among women with low PAH exposure. Further, using the linear-exponential model for smoking EOR we found that the difference in smoking effects between the wild type and the variant resulted from the effects of cigarettes per day and not pack-years. This finding suggests that a dose-response mechanism and a saturation effect might underlie the smoking-mediated association between CYP1A1 and lung cancer risk. The gene expression analysis supported this finding. In fact, the lower expression of CYP1A1 among never smokers and higher expression among current smokers in association with the SNPs at chr15q24.1 was consistent with the observed protective effect for lung cancer among never smokers and risk among smokers in association with variants in CYP1A1/A2.

Our data also showed that the minor allele of CYP1B1 rs10175368 was significantly protective for adenocarcinoma of the lung (OR = 0.80, 95%CI = 0.69–0.93) and a similar protective effect was observed for the minor allele of CYP1B1 rs9341266 (r² = 0.30), as well as for the cumulative sum of the two minor alleles. In addition, according to the HapMap database, CYP1B1
gene expression analysis supported this finding. In fact, the lower expression of CYP1A1 among never smokers and higher expression among current smokers in association with the SNPs at chr15q24.1 was consistent with the observed protective effect for lung cancer among never smokers and risk among smokers in association with variants in CYP1A1/A2.

Table 5. False positive report probability.

| Gene/SNP          | MAF | Controls | Cases | Test | OR(**) | P-value | Power(*) | Prior Probabilities |
|-------------------|-----|----------|-------|------|--------|---------|----------|---------------------|
|                   |     |          |       |      |        |         |          |                     |
| CYP2A5/rs1801272  | 0.041| 2019     | 1859  | (a)  | 0.720  | 0.026   | 0.772    | 0.033   | 0.092 | 0.233 | 0.769 | 0.971 |
| CYP1B1/rs2934568  | 0.042| 2015     | 767   | (b)  | 1.380  | 0.032   | 0.632    | 0.048   | 0.132 | 0.313 | 0.834 | 0.981 |
| CYP1B1/rs9341266  | 0.060| 2032     | 774   | (b)  | 0.741  | 0.046   | 0.609    | 0.070   | 0.185 | 0.405 | 0.882 | 0.987 |
| CYP1B1/rs162556   | 0.446| 2023     | 768   | (b)  | 1.156  | 0.031   | 0.673    | 0.044   | 0.121 | 0.293 | 0.820 | 0.979 |
| CYP1B1/rs10175368 | 0.282| 2022     | 772   | (b)  | 0.799  | 0.003   | 0.907    | 0.003   | 0.010 | 0.029 | 0.247 | 0.768 |
| CYP2A6/rs1801272  | 0.041| 2019     | 481   | (c)  | 0.467  | 0.007   | 0.926    | 0.008   | 0.022 | 0.064 | 0.428 | 0.883 |
| CYP1A1/rs2606345  | 0.358| 2028     | 1866  | (d)  | 0.687  | 0.005   | 0.795    | 0.006   | 0.019 | 0.054 | 0.384 | 0.863 |
| CYP1A2/rs11072508 | 0.388| 2031     | 1861  | (d)  | 0.822  | 0.038   | 0.545    | 0.065   | 0.173 | 0.386 | 0.873 | 0.986 |
| CYP1A2/rs4886410  | 0.383| 2030     | 1866  | (d)  | 0.829  | 0.047   | 0.511    | 0.084   | 0.216 | 0.453 | 0.901 | 0.989 |
| CYP2A6/rs1801272  | 0.041| 2019     | 1859  | (d)  | 1.508  | 0.026   | 0.757    | 0.033   | 0.093 | 0.236 | 0.773 | 0.972 |
| CYP1A1/rs2606345  | 0.358| 2028     | 774   | (e)  | 0.739  | 0.022   | 0.481    | 0.044   | 0.121 | 0.291 | 0.819 | 0.979 |

FPRP values for the nominally significant (p-value < 0.05) results from test of main single SNP effects (Table 3) and of SNP-smoking interaction effects (Table 4). FPRP is computed according to the formula \( \text{FPRP} = \frac{1}{n \cdot \text{P-value}} \), where \( n \) is the number of SNPs and the P-value represents the Prior Probability ranging from 0.01 to 0.5. FPRP values less than 0.2 are in italic, FPRP values between 0.2 and 0.5 are bold, and FPRP values larger than 0.5 are the rest.

(a) Test for main genetic effect among all subjects.
(b) Test for main genetic effect among controls and adenocarcinoma cases.
(c) Test for main genetic effect among controls and squamous carcinoma cases.
(d) Test for gene-smoking interaction among all subjects.
(e) Test for gene-smoking interaction among controls and adenocarcinoma cases.

(*) OR indicates the measured odds ratio for the main genetic effect for tests (a), (b), and (c), and the measured odds-ratio ratio for the gene-smoking interaction effect for tests (d) and (e).

(**) The statistical power to detect the measured OR with a significance level of 0.05 was computed by means of the QUANTO software (http://hydra.usc.edu/gxe).
doio:10.1371/journal.pone.0005652.0005

Two haplotypes in EPHX1 compared to all other haplotypes were significantly associated with lung cancer in the overall population and in adenocarcinoma cases only: TGCCACCT as a risk factor and CGGCGCCT as a protective factor. In addition, we identified a borderline significant protective association with a three-locus haplotype which was also contained in the 8SNP haplotype and was present in approximately 3% of the population. These findings suggest that more than a hundred people in our study carried a three-variant haplotype resulting in a decreased lung cancer risk. The protective effect was even stronger for the smaller number of subjects (1%) who carried a combination of these three SNPs and the remaining 5 SNPs in the 8-locus haplotype. Since the significant associations with lung cancer were based on relatively rare haplotypes, replication will be needed in order to validate this finding. None of the 8 SNPs was significantly associated with lung cancer in the overall population when analyzed separately. This result, if confirmed, demonstrates that the effect of multiple SNPs on lung cancer may be important even if most individual SNPs do not show significant association. This may explain why previously published results, which are based on a limited number of EPHX1 polymorphisms, were inconsistent. In particular, EPHX1 rs2234922 has been previously associated both with risk [39] and with protection [57] for lung cancer. This SNP was not associated with lung cancer in our data. Nevertheless it was one of the three SNPs that differentiate the two significant haplotypes reported here. The other two SNPs were EPHX1 rs1031741, in medium LD with EPHX1 rs2234922, and EPHX1 rs2292568, nominally significantly associated with risk of lung adenocarcinoma in our data (see Table 3). We did not find a significant association between EPHX1 polymorphisms and gene expression. Measurements of epoxide hydrolase activity in lung cancer patients carrying these haplotypes will be needed in order to understand the biological mechanism that underlies this finding.
rs10175368 is in LD with 4 other SNPs in the same chromosomal region (rs2551190, rs4646430, rs4646429, and rs10175338, see Supplemental Figure S1B). These 4 SNPs are likely to be characterized by the same protective association. Previous results on CYP1B1 polymorphisms and lung cancer have been limited to the four non-synonymous SNPs rs10012, rs1056027, rs1056036 and rs1000440 [27–30,59,60]. None of the reported positive findings have been consistently replicated, except for rs10012, associated with lung cancer risk in two independent studies [28,30]. Our data on rs1800440 did not show any significant association with lung cancer. The three other non-synonymous SNPs were not evaluated in the current study. However, our SNPs were selected with an attempt to cover other regions of the gene. According to our data, variants other than those in the coding region could alter lung cancer risk. Polymorphisms in CYP1B1 have been associated with decreased PAH metabolism [26]. The significant protective effect of the CYP1B1 rs10175368 variant allele could be due to a lower level of smoking carcinogens in subjects carrying the variant allele. We did not find a significant effect on CYP1B1 gene expression for the two SNPs in CYP1B1 associated with a protection for adenocarcinoma. However, when we considered all seven polymorphisms in CYP1B1 together and studied their effect on gene expression, we found a significant increase in CYP1B1 gene expression among current smokers. The CYP1B1 gene is known to be highly expressed in lung tissues of lung cancer patients. Our result supports previous findings of CYP1B1 gene over-expression among current smokers [61] and suggests a possible involvement of CYP1B1 polymorphisms as a mechanism for differential expression.

The CYP2A6 rs1801272 polymorphism, which results in an amino acid change from Leu to His, was significantly associated with a decreased risk for squamous cell carcinoma, a strictly smoking-related malignancy. Interestingly, the same SNP was associated with a decrease in cigarettes per day in controls, confirming a previously hypothesized role of this gene in tobacco smoking addiction [46]. Our report provides the first confirmation of this finding in a population-based sample. In addition, the A allele of CYP2A6 rs1801272 showed a significant protective effect in ever smokers but no effect in never smokers, with a nominally significant genotype-smoking interaction due to the effect of cigarettes per day and not pack-years. The CYP2A6 gene is characterized by multiple polymorphisms and genomic repetitive elements in the regulatory regions, which make a complete coverage of the gene extremely challenging. Moreover, most variants are very rare in the general population and would not be identifiable even in a large sample size as ours. We genotyped CYP2A6 rs1801272 (also known as CYP2A6*2) because this SNP is relatively common (4% in our population), has been well characterized in previous functional studies [46], and showed controversial associations with cancer and smoking dependence [43–46]. Our findings of an association with both lung cancer risk and tobacco addiction warrant further investigation based on a more complete coverage of this gene.

The size of our population provides unusual power for confirming previously reported associations. Our data do not support proposed associations between lung cancer and EPHX1 rs2234922, CYP1B1 rs1800440, and MPO rs2335227. The confidence in our significant results was supported by the low FPRP values (see Table 5) observed for prior probabilities of 0.10 or more given the strong prior probabilities of the selected phase I genes being involved in lung cancer risk.

At the time that this project was initiated, there was less genotype data available with which to select SNPs to cover haplotype blocks. Nevertheless, based on the existing SNP500Can-
A. Main effect of genotype. The main effect of the variant genotypes on the risk of lung cancer was estimated by odds ratios and their 95% confidence intervals using unconditional logistic regression analysis. Homozygosity for the more frequent allele and their 95% confidence intervals using unconditional logistic regression: Logit(LC) = \alpha + \beta \times SNPs + \gamma \times covariates (1)

where \( \alpha \) is the intercept, \( \beta \) is the regression coefficient for the cumulative number of pack-years, and \( \gamma \) is the regression coefficient for the categorical variable. The main effect of the variant was derived from a polytomous logistic regression defined the standard Wald chi-square test statistic using the small cell carcinoma of cases. In the analysis by histology, we performed stratified analyses by smoking status (never/ever) of smoking dose (pack-years), smoking intensity (cigarettes per day), categorical variable. Age, sex, geographical location, cumulative continuous in logit scale in the model, and the categorical analysis effect of SNP was conducted by including the SNP variable as significance using two-sided Wald tests. The trend test for the regression analysis. Homozygosity for the more frequent allele and their 95% confidence intervals using unconditional logistic genotypes on the risk of lung cancer was estimated by odds ratios

B. Genotype-smoking interaction. We evaluated genotype-smoking interactions using a likelihood ratio test to compare the following two models:

i. Logit(LC) = \alpha_0 + \beta_1 \times SNPs_{g(n)}\times smoking status + \gamma \times covariates

ii. Logit(LC) = \alpha_0 + \beta_1 \times SNPs_{g(n)}\times smoking status + 0 \times SNPs \times smoking status + \gamma \times covariates

For polymorphisms showing the presence of a genotype-smoking interaction in the association with lung cancer, we fitted a model for the excess odds ratio of smoking (EOR) \([54]\) in order to separate the contribution of total exposure and intensity in the interaction with the polymorphism. Specifically, we fitted the following 3-parameter linear-exponential model which described the OR in terms of continuous pack-years (\(d\)) and continuous cigarettes per day (\(n\)): OR(d,n) = 1 + \beta \times d + g(n), where \(\beta\) is the EOR at \(g(n) = 1\), i.e., the EOR/pack-year, and \(g(.)\) is a function that describes the influence of changing cigarettes per day on the strength of the lung cancer and pack-years association. Based on an empirical evaluation, we used a two parameter form for \(g(.)\), where \(g(n) = \exp\{\psi_1 \times n + \psi_2 \times n^2\}\). The component, \(\beta \times g(n)\), describes the EOR per pack-year and its variation with cigarettes per day and thus the influence of the delivery rate, i.e., increasing cigarettes per day and decreasing duration of exposure. We expanded this model to incorporate genotype (\(s\), where \(s = 1\) and \(0\) denote the variant and wild type forms, respectively), using: OR\((s,d,n) = \exp\{\psi(s) \times [1 + \beta \times d + g(n)]\}\), where the subscripts denote separate parameters for each genotype. We fitted the model to data on never and current smokers (including subjects who quit smoking less than two years before the study) and on never and former smokers (subjects who quit smoking more than two years before the study), and used likelihood ratio tests to compare homogeneity of the effects of pack-years, i.e., \(\beta_1 = \beta_0\), and/or smoking intensity, i.e., \(\psi_1 = \psi_0\), and \(\psi_2 = \psi_2\). C. Joint SNPs. We analyzed multiple SNPs jointly to test whether the overall lung cancer risk was determined by the combination of multiple SNPs within the same gene and/or of multiple genes within the same pathway, even if each SNP may have had only a modest effect size individually.

c1. Under the assumption that the effect on lung cancer of each SNP was cumulative, we implemented the following model:

Logit(LC) = \alpha + \beta \times \sum_{k}^{n} (SNP_k) + \gamma \times covariates

where \(k = 1, \ldots, n\) represents a collection of SNPs belonging to the same gene or a collection of SNPs belonging to genes in the same pathway (e.g. phase I, \(n = 25\) i.e. all SNPs were grouped together). \(SNP_1 = 0\) for the homozygote most common allele, \(SNP_2 = 1\) for the heterozygote allele, and \(SNP_3 = 2\) for the homozygote minor allele. \(\beta\) is the regression coefficient for the cumulative number of variants \(\Sigma^s \times (SNP_k)\). We estimated the overall risk of lung cancer associated with each selected group of \(n\) SNPs by computing \(OR = \exp(\beta)\) in the overall population, in never smokers, and in ever smokers separately. We estimated smoking-genotype interaction using the likelihood ratio test. Note that in this model we do not assume nor infer a risk direction for each minor allele. This approach will be powerful if minor alleles for all SNPs have effects in the same direction, but there may be loss of power if minor alleles for some SNPs affect lung cancer risk in opposite directions and their contribution to the overall risk cancels with each other.

c2. For all genes represented in our data by two or more SNPs, we computed pairwise linkage disequilibrium (LD) using the Haplovew software and carried out haplotype analysis using the haplo.stats R-package.

D. Gene Expression. We evaluated to the extent possible, the effect of polymorphisms \(SNP^{G}\) from a given gene \(G\) on the gene expression of the same gene \(G\), and specifically the effect related to lung cancer. We first estimated the overall effect of each group of SNPs \((SNP^{G})\) on lung cancer according to the additive model

Logit(LC) = \alpha_0 + \sum_{k}^{n} (\beta_k \times SNP^{G}_k) + \gamma \times covariates

where \(\beta_k\) are the regression coefficients for the \(n\) SNPs in \(G\). Second, we used the \(\beta_k\) estimated from equation (2) to compute the overall effect of each group of polymorphisms \(SNP^{G}\) on the change of gene expression of \(G\) (Exp\(_G\)) by solving the following logistic regression:

Exp\(_G\) = \alpha_1 + \delta \times \sum_{k}^{n} (\beta_k \times SNP^{G}_k).

(3)
According to equation (3), $\delta > 0$ indicates an increase and $\delta < 0$ a decrease in the gene expression of the gene $G$, due to the overall effect of the polymorphisms $\Delta V_1 \Delta V_2$ on lung cancer. Basically, we used the SNPs regression score for lung cancer and verified whether it was positively or negatively associated with gene expression in non tumor tissue samples from a subgroup of cases. Note that since we lack gene expression data from healthy controls because no fresh frozen lung tissue samples can be collected from healthy people, we cannot measure directly the association between gene expression and lung cancer risk. Combining equation (2) and (3) instead, we are able to obtain such information. The described gene expression analysis was performed overall and, separately, among never smokers, former smokers, and current smokers.

E. Multiple testing and a priori knowledge considerations. We considered significant those results with a $p$-value less than (or equal to) 0.01. This choice was a compromise between a more stringent Bonferroni-corrected $p$-value and the loss in power from getting the threshold for significance too low. In addition, we referred to results with $p$-value between 0.01 and 0.05 as nominally significant, and considered them as notable when consistent across different analyses. Given the number of tested hypotheses in the single SNP analyses (25 tests corresponding to the 25 SNPs for the single SNP analysis and 5 tests when SNPs were grouped by genes) we took multiple testing into account. Our approach to multiple testing was informed by the selection strategy for the Phase I genes selected. Of note, each of the genes included has substantial mechanistic and at least some population data which support an association with lung cancer, as we have described in the introduction. We recognize that quantifying this a priori knowledge for each SNP is challenging, because of the heterogeneity of results in the literature and because most results actually refer to genes and not to our specific SNPs. In order to incorporate the effect of both multiple testing and a priori knowledge considerations, we computed the False Positive Report Probability (FPRP) [55] to characterize the noteworthyness for all the significant and nominally significant results from single SNP analyses for a range of prior probabilities. The statistical power to detect the measured OR given a type I error rate of 0.05 was computed by means of the QUANTO software (http://hydra.usc.edu/gxe).

Supporting Information

Figure S1  Genes coverage for EPHX1, CYP1B1 and CYP1A1/ A2
Found at: doi:10.1371/journal.pone.0005652.s001 (0.68 MB DOC)

Figure S2  Results for linkage disequilibrium and haplotype analyses in CYP1A1 and CYP1B1.
Found at: doi:10.1371/journal.pone.0005652.s002 (0.04 MB DOC)

Table S1  Associations between SNPs and lung cancer overall and major histology subtypes.
Found at: doi:10.1371/journal.pone.0005652.s003 (0.79 MB DOC)

Table S2  Joint SNP analysis.
Found at: doi:10.1371/journal.pone.0005652.s004 (0.05 MB DOC)

Table S3  Haplotype and three-marker moving window analyses in EPHX1.
Found at: doi:10.1371/journal.pone.0005652.s005 (0.14 MB DOC)

Table S4  Gene expression and SNP correlation analysis.
Found at: doi:10.1371/journal.pone.0005652.s006 (0.04 MB DOC)

Text S1  Results for linkage disequilibrium and haplotype analyses in EPHX1 and CYP1B1.
Found at: doi:10.1371/journal.pone.0005652.s007 (0.03 MB DOC)

Acknowledgments

We are deeply indebted to the EAGLE participants and the study collaborators (listed in http://dceg.cancer.gov/eagle).

Author Contributions

Conceived and designed the experiments: DC ACP AMG LG SW PAB MT NEC AWB MTL. Performed the experiments: RW LB SJC. Analyzed the data: MR KY JL SW NC AWB. Contributed reagents/materials/analysis tools: DC ACP PAB MT NEC MTL. Wrote the paper: MR MTL. Reviewed the manuscript: KY JL DC AMG LG SW NC AWB.

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, et al. (2008) Cancer statistics 2008. CA Cancer J Clin 58: 71-96.
2. Wright GS, Gaudet ME. (2000) Early detection and prevention of lung cancer. Curr Oncol 12: 143-148.
3. Shields PG (1999) Molecular epidemiology of lung cancer. Ann Oncol 10: S7–11.
4. Greiner P, Gromadzinska J, Wasosicw W. (2007) Polymorphism of selected enzymes involved in detoxification and biotransformation in relation to lung cancer. Lung Cancer 57: 1–25.
5. Kiyohara C, Yoshinaka K, Takayama K, Nakanishi Y (2007) Lung cancer susceptibility: are we on our way to identifying a high-risk group? Future Oncol 3: 617-627.
6. Landi MT, Consonni D, Rotunno M, Bergen AW, Goldstein AM, et al. (2008) Environment And Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. BMC Public Health 8: 203.
7. Hoffmann D, Hoffmann I, El-Rayyesy K. (2001) The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. Chem Res Toxicol 14: 767-790.
8. Shimada T, Fuji-Kuriyama Y (2004) Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1. Cancer Sci 95: 1-6.
9. Kim JH, Stansbury KH, Walker NJ, Trush MA, Strickland PT, Sutter TR (1998) Metabolism of benz[a]pyrene and benzo[a]pyrene-7,8-diol by human cytochrome P450 1B1. Carcinogenesis 19: 1047-1053.
10. Spinke DC, Spinke BC, Cao JQ, DePasquale JA, Pentecost BT, et al. (1998) Differential expression of CYP1A1 and CYP1B1 in human breast epithelial cells and breast tumor cells. Carcinogenesis 19: 291-298.
11. Spinack SD, Hurteau GJ, Fasico MJ, Kaminsky LS (2003) Phase I and II carciogen metabolism gene expression in human lung tissue and tumors. Clin Cancer Res 9: 6002-6011.
12. Nebert DW, Dahlus TP, Okey AB, Gonzalez FJ (2004) Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem 279: 23847-23850.
13. Tefre T, Ryberg D, Haugen A, Nebert DW, Skau V, et al. (1993) Human CYP1A1 (cytochrome P1/450) 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.
14. Cascorbi I, Brockmoller J, Roots I (1996) A C1048T A polymorphism in exon 7 of human CYP1A1: population frequency, mutation linkages, and impact on lung cancer susceptibility. Cancer Res 56: 4963-4969.
15. Houlston RS (2000) CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. Pharmacogenetics 10: 105-114.
16. Le ML, Guo C, Bingham S, Bouchard C, Cascorbi I, et al. (2003) Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States). Cancer Causes Control 14: 339-346.
17. Shah PP, Singh AP, Singh M, Mathur N, Pant MC, et al. (2008) Interaction of cytochrome P4501A1 genotypes with other risk factors and susceptibility to lung cancer. Mutat Res 639: 1-10.
10. Alexandrov K, Cascorbi I, Rojas M, Bouvier G, Krick E, Bartsch H (2002) CYP1A1 and GSTM1 genotypes affect benzo[a]pyrene DNA adducts in smokers’ lung: comparison with aromatic/hydrophobic adduct formation. Carcinogenesis 23: 1890–1897.

19. Ueda R, Betkai H, Nagaia K, Kimura S, Gonzalez FJ, et al. (2006) A common regulatory region functions bidirectionally in transcriptional activation of the human CYP1A1 and CYP1A2 genes. Mol Pharmacol 69: 1924–1930.

30. Zienolddiny S, Campa D, Lind H, Ryberg D, Skaug V, et al. (2008) A novel promoter polymorphism is associated with lung cancer risk. Cancer Res 68: 5598–5604.

32. Fretland AJ, Omiecinski CJ (2000) Epoxide hydrolases: biochemistry and regulation. Biochim Biophys Acta 695: 251–270.

34. Coller JK, Fritz P, Zanger UM, Siegle I, Eichelbaum M, et al. (2001) Genetic regulation and function of CYP2A6 in primary human lung cancer. Cancer Res 61: 4541–4546.

35. Hassett C, Aicher L, Sidhu JS, Omiecinski CJ (1994) Human microsomal epoxide hydrolase: properties, chemical study in normal tissues, and benign and malignant tumours. Histochem J 33: 329–336.

36. Raaka S, Hassett C, Boffetta P, Ashfield M, et al. (2000) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit on 15q25. Nature 452: 633–637.

37. Spitz MR, Amos CI, Dong Q, Lin J, Wu X (2008) The CHRNA5-A3 region on chromosome 15q24–25.1 is a risk factor both for nicotine dependence and for lung cancer. J Natl Cancer Inst 100: 1552–1556.

38. Thorgeirsson TE, Geller S, Sulem P, Rafnar T, Wiste A, et al. (2000) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452: 638–642.

40. Voho A, Metsola K, Anttila S, Impivaara O, Jarvisalo J, et al. (2006) EPHX1 polymorphism of CYP2A6 as one of the potential determinants of tobacco-related cancer risk. Biochem Biophys Res Commun 338: 306–310.

41. Lewis DF, Dickins M, Lake BG, Eddershaw PJ, Tarbit MH, Goldfarb PS (1999) CYP1A1 and CYP1A2 gene expression in vitro of human lung cancer. Lung Cancer 23: 1969–1977.

42. Lewis DF, Dickins M, Lake BG, Eddershaw PJ, Tarbit MH, Goldfarb PS (1999) Molecular modelling of the human cytochrome P450 isomorph CYP2A6 and investigations of CYP2A6 substrate selectivity. Toxicology 133: 1–33.

43. London SJ, Idle JR, Daly AK, Coe-Tyre GA (1999) Genetic variation of CYP2A6, smoking, and risk of cancer. Cancer 33: 880–889.

44. Ariyoshi N, Miyamoto M, Umemoto Y, Kunitoh H, Aka-Akita H, et al. (2002) Genetic polymorphism of CYP2A6 gene and tobacco-induced lung cancer risk in male smokers. Cancer Epidemiol Biomarkers Prev 11: 890–894.

45. Kamatani T, Fujieda M, Koyanagi K, Iwano S, Kunitoh H (2003) Genetic polymorphism of CYP2A6 as one of the potential determinants of tobacco-related cancer risk. Biochem Biophys Res Commun 308: 306–310.

46. Pianezza ML, Sellers EM, Tyndale RF (1998) Nicotine metabolism defect reduces smoking. Nature 393: 750.

47. Schmekel B, Severs L, Xu SY, Venge P (2000) Human neutrophil lipocalcin (HNL) and myeloperoxidase (MPO). Studies of lung lavage fluid and lung tissue. Respir Med 94: 564–568.

48. Mallet WG, Mosebroek DR, Truth MA (1991) Activation of (+)-trans-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene to diol epoxides by human polymorphonuclear leukocytes or myeloperoxidase. Carcinogenesis 12: 521–524.

49. Kadlubar FF, Butler MA, Kaderlik KR, Chou HC, Lang NP (1992) Polymorphisms for aromatic amine metabolism in humans: relevance for human carcinogenesis. Environ Health Perspect 98: 69–74.

50. Truth MA, Severs L, Kender TW (1985) Oxidant-dependent metabolic activation of polycyclic aromatic hydrocarbons by phorbol ester-stimulated human polymorphonuclear leukocytes: possible link between inflammation and cancer. Proc Natl Acad Sci U S A 82: 5194–5198.

51. Van Schooten FJ, Boots AW, Knaapen AM, Godschalk RW, Maas LM, et al. (2004) Myeloperoxidase (MPO)-463G>R reduces MPO activity and DNA adduct levels in bronchoalveolar lavages of smokers. Cancer Epidemiol Biomarkers Prev 13: 828–833.

52. Feyler A, Voho A, Bouchard C, Kuskanen K, Dayer P, et al. (2002) Point: a polymorphism and lung cancer risk. Cancer Epidemiol Biomarkers Prev 11: 1550–1554.

53. Xu LL, Liu G, Miller DP, Zhou W, Lynch TJ, et al. (2002) Counterpoint: the myeloperoxidase -463G>R polymorphism does not decrease lung cancer susceptibility in Caucasians. Cancer Epidemiol Biomarkers Prev 11: 1553–1559.

54. Lubin JH, Kogevinas M, Silverman D, Malats N, Garcia-Closas M, et al. (2007) Evidence for an intensity-dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. Int J Epidemiol 36: 236–241.

55. Wacholder S, Chanock S, Garcia-Closas M, El GI, Rothman N (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 96: 434–442.

56. Schottenfeld D, Fraumeni F (1996) Cancer Epidemiology and Prevention. Oxford University Press, New York, 2nd ed., 561 pp.

57. To-Figueras J, Gene M, Gomez-Catalan J, Pique E, Borrego N, Corbella J (2006) Comprehensive evaluation of variability in nicotine metabolism and CYP2A6 subunit genes on 15q25. Nature 452: 633–637.

58. Schmekel B, Severs L, Xu SY, Venge P (2000) Human neutrophil lipocalcin (HNL) and myeloperoxidase (MPO). Studies of lung lavage fluid and lung tissue. Respir Med 94: 564–568.

59. Mallet WG, Mosebroek DR, Truth MA (1991) Activation of (+)-trans-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene to diol epoxides by human polymorphonuclear leukocytes or myeloperoxidase. Carcinogenesis 12: 521–524.

60. Van Schooten FJ, Boots AW, Knaapen AM, Godschalk RW, Maas LM, et al. (2004) Myeloperoxidase (MPO)-463G>R reduces MPO activity and DNA adduct levels in bronchoalveolar lavages of smokers. Cancer Epidemiol Biomarkers Prev 13: 828–833.

61. Feyler A, Voho A, Bouchard C, Kuskanen K, Dayer P, et al. (2002) Point: a polymorphism and lung cancer risk. Cancer Epidemiol Biomarkers Prev 11: 1550–1554.

62. Xu LL, Liu G, Miller DP, Zhou W, Lynch TJ, et al. (2002) Counterpoint: the myeloperoxidase -463G>R polymorphism does not decrease lung cancer susceptibility in Caucasians. Cancer Epidemiol Biomarkers Prev 11: 1553–1559.

63. Lubin JH, Kogevinas M, Silverman D, Malats N, Garcia-Closas M, et al. (2007) Evidence for an intensity-dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. Int J Epidemiol 36: 236–241.

64. Wacholder S, Chanock S, Garcia-Closas M, El GI, Rothman N (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 96: 434–442.

65. Schottenfeld D, Fraumeni F (1996) Cancer Epidemiology and Prevention. Oxford University Press, New York, 2nd ed., 561 pp.