REVIEW

Association Between Genetic Variants on Chromosome 15q25 Locus and Objective Measures of Tobacco Exposure

Marcus R. Munafò, Maria N. Timofeeva, Richard W. Morris, David Prieto-Merino, Naveed Sattar, Paul Brennan, Elaine C. Johnstone, Caroline Relton, Paul C. D. Johnson, Donna Walther, Peter H. Whincup, Juan P. Casas, George R. Uhl, Paolo Vineis, Sandosh Padmanabhan, Barbara J. Jefferis, Antoinette Amuzu, Elio Riboli, Mark N. Upton, Paul Aveyard, Shah Ebrahim, Aroon D. Hingorani, Graham Watt, Tom M. Palmer, Nicholas J. Timpson, EPIC Study Group, George Davey Smith

Correspondence to: Marcus R. Munafò, PhD, School of Experimental Psychology, University of Bristol, 12a Priory Rd, Bristol BS8 1TU, UK (e-mail: marcus.munafo@bristol.ac.uk).

Background	Two single-nucleotide polymorphisms, rs1051730 and rs16969968, located within the nicotinic acetylcholine receptor gene cluster on chromosome 15q25 locus, are associated with heaviness of smoking, risk for lung cancer, and other smoking-related health outcomes. Previous studies have typically relied on self-reported smoking behavior, which may not fully capture interindividual variation in tobacco exposure.
Methods	We investigated the association of rs1051730 and rs16969968 genotype (referred to as rs1051730–rs16969968, because these are in perfect linkage disequilibrium and interchangeable) with both self-reported daily cigarette consumption and biochemically measured plasma or serum cotinine levels among cigarette smokers. Summary estimates and descriptive statistical data for 12 364 subjects were obtained from six independent studies, and 2932 smokers were included in the analyses. Linear regression was used to calculate the per-allele association of rs1051730–rs16969968 genotype with cigarette consumption and cotinine levels in current smokers for each study. Meta-analysis of per-allele associations was conducted using a random effects method. The likely resulting association between genotype and lung cancer risk was assessed using published data on the association between cotinine levels and lung cancer risk. All statistical tests were two-sided.
Results	Pooled per-allele associations showed that current smokers with one or two copies of the rs1051730–rs16969968 risk allele had increased self-reported cigarette consumption (mean increase in unadjusted number of cigarettes per day per allele = 1.0 cigarette, 95% confidence interval [CI] = 0.57 to 1.43 cigarettes, $P = 5.22 \times 10^{-6}$) and cotinine levels (mean increase in unadjusted cotinine levels per allele = 138.72 nmol/L, 95% CI = 97.91 to 179.53 nmol/L, $P = 2.71 \times 10^{-11}$). The increase in cotinine levels indicated an increased risk of lung cancer with each additional copy of the rs1051730–rs16969968 risk allele (per-allele odds ratio = 1.31, 95% CI = 1.21 to 1.42).
Conclusions	Our data show a stronger association of rs1051730–rs16969968 genotype with objective measures of tobacco exposure compared with self-reported cigarette consumption. The association of these variants with lung cancer risk is likely to be mediated largely, if not wholly, via tobacco exposure.

J Natl Cancer Inst 2012;104:740–748

Genome-wide association studies (GWAS) have identified specific common genetic variants within a region on the long arm of chromosome 15 (15q25) as risk factors for lung cancer (1–3) and other smoking-related health outcomes, such as chronic obstructive pulmonary disease (4,5), peripheral arterial disease (3,6), low birth weight in offspring (7), and lower body mass index (8). This region contains the α5, α3, and β4 nicotinic acetylcholine receptor subunit gene cluster, *CHRNA5-CHRNA3-CHRNB4* (*CHRNA5-A3-B4*); although most studies to date have focused on the rs16969968 and rs1051730 single-nucleotide polymorphisms (SNPs) within this gene cluster, there is emerging evidence that other SNPs within this region are also associated with smoking behavior (9). A few studies have suggested that rs16969968 and rs1051730 SNPs do not relate to smoking history (2,10), but most

have found them to be associated with a variety of smoking-related phenotypes, including tobacco dependence, heaviness of smoking (ie, number of cigarettes per day), risk of relapse after quitting smoking, age at initiation of smoking, and subjective response to the first cigarette smoked (1,3,7,9,11–22).

The first published study that investigated this region in relation to cigarette smoking identified the rs16969968 SNP (23), and this was subsequently confirmed at a genome-wide statistically significant threshold (24). A recent study showed that the minor allele of the rs16969968 missense polymorphism, D398N, in *CHRNA5* was associated with a reduced response to a nicotinic agonist in vitro (25) and may therefore be the functional variant responsible for the association with smoking quantity. The rs1051730 SNP is located within the *CHRNA3* gene, in a linkage

disequilibrium (LD) block within *CHRNA5-A3-B4*, and has also been shown to be associated with heaviness of smoking (3). This SNP is in perfect linkage disequilibrium with rs16969968 in samples of European ancestry (LD decay [D′] = 1.0, LD correlation coefficient [R^2] = 1.0), and the two SNPs are commonly treated interchangeably (hereafter referred to as rs1051730–rs16969968). Across these studies, minor allele (rs1051730 T, rs16969968 A) carriers show increased risk of heavier smoking, as well as lung cancer and other health-related outcomes, compared with wild-type (rs1051730 C, rs16969968 G) homozygotes.

However, these studies typically rely on self-reported measures of smoking behavior, which do not fully capture interindividual variation in tobacco exposure (26). Two small studies have reported on the association of rs1051730–rs16969968 with heaviness of smoking among current smokers, measured by self-reported daily cigarette consumption and by levels of cotinine and other nicotine metabolites in serum (27,28). In one study, several nicotine metabolites in the urine were used to produce an index of nicotine equivalents (28). It showed that the risk alleles for lung cancer were associated with higher urinary nicotine equivalents among smokers, and this association remained after adjustment for self-reported daily cigarette consumption. This suggests that other aspects of smoking behavior that influence exposure, such as depth of inhalation, are associated with chromosome 15 risk alleles for lung cancer.

Misreporting of smoking behavior by smokers (eg, reporting that they smoke fewer cigarettes than they actually do) will also reduce the reliability of self-reported measures as an index of exposure. Therefore, the use of self-report measures of smoking behavior could lead to relationships between risk alleles and disease outcomes such as lung cancer, apparently independently of smoking intensity. Consequently, this would imply a direct association of genotype with risk of disease outcomes, when, in fact, the association may be entirely because of tobacco exposure. If this is the case, the *CHRNA5-A3-B4* SNPs should be more strongly associated with objective measures of tobacco exposure than with self-report measures. This prediction is supported by another small genetic association study (27), which investigated serum cotinine levels and self-reported smoking, and showed a considerably stronger association with the objective measures of exposure (ie, cotinine levels).

Two previous studies (27,28) that examined the association between *CHRNA5-A3-B4* SNPs and both self-reported and objective measures of tobacco exposure were small. In this study, we extended this preliminary work by conducting a much larger collaborative investigation. We tested the association of rs1051730–rs16969968 with both self-reported daily cigarette consumption and cotinine levels, which allowed us to estimate the extent to which the association of these SNPs with smoking behavior may have been underestimated, and thus the smoking-mediated influence of these SNPs on lung cancer and other conditions underappreciated.

Materials and Methods

Study Population

Data on self-reported smoking status (never smoker, former smoker, and current smoker), genotype (rs16969968 or rs1051730),

CONTEXT AND CAVEATS

Prior knowledge
Two interchangeable single-nucleotide polymorphisms, rs1051730 and rs16969968, are associated with heaviness of smoking, risk for lung cancer, and other smoking-related health outcomes. Previous studies have mostly relied on retrospective self-reported smoking behavior, which may have produced underestimated associations and masked the contribution of smoking to the observed association of these polymorphisms with lung cancer and other health outcomes.

Study design
Data from six independent studies were used to assess associations of rs1051730 and rs16969968 genotype (referred to as rs1051730–rs16969968) with self-reported daily cigarette consumption and plasma or serum cotinine levels among cigarette smokers, and meta-analysis of per-allele effects was conducted. The likely resulting association between genotype and lung cancer risk was also assessed.

Contribution
Associations of rs1051730–rs16969968 risk allele with increased self-reported cigarette consumption per day and increased cotinine levels were observed; however, the association with cotinine level was much stronger. This increase in cotinine levels indicated an increased risk of lung cancer (per-allele odds ratio = 1.31).

Implication
The association of rs1051730–rs16969968 genotype with lung cancer risk is mediated to a large extent, if not completely, by tobacco exposure.

Limitation
Analysis was based on current smoking status and not on lifetime exposure and also did not adjust for factors known to influence nicotine metabolism.

From the Editors

and cotinine levels determined in blood (plasma or serum), as well as cigarette consumption (cigarettes per day) among current smokers, were available from participants of self-reported European ancestry in six independent studies (British Regional Heart Study [BRHS], n = 385 subjects; British Women's Heart and Health Study [BWHHS], n = 400 subjects; European Prospective Investigation into Cancer and Nutrition [EPIC], n = 759 subjects; Midspan, n = 499 subjects; Patch II, n = 451 subjects; Patch in Practice [PiP], n = 438 subjects) that contributed to this analysis (2,29–40). All studies received appropriate ethics approval, and informed consent was obtained from each participant. These studies are described in detail in Supplementary Methods (available online).

The data from the EPIC study were drawn from a case–control lung cancer study nested within the EPIC cohort. Blood and questionnaire data on self-reported smoking behavior were collected in a prospective manner as part of the recruitment procedure and before cancer diagnosis in case subjects (41). However, cotinine levels and self-reported cigarette consumption in current smokers differed between case subjects (EPIC Case) and control subjects (EPIC Control). Mean cotinine levels were higher in the EPIC

Case subsample (1391 nmol/L, SD = 588 nmol/L) than the EPIC Control subsample (1006 nmol/L, SD = 589 nmol/L) even after adjustment for cigarette consumption (two-sided Kruskal–Wallis $P < .001$), indicating possible differences in smoking behavior (eg, depth of inhalation). This difference at baseline may have contributed to the likelihood of a subsequent lung cancer diagnosis. Therefore, the EPIC case subjects and control subjects were analyzed and presented separately in this study.

Statistical Analysis

In each contributing study, linear regression was used to calculate per-allele associations of rs1051730–rs16969968 genotype on daily cigarette consumption (cigarettes per day) and cotinine levels (nmol/L) in current smokers. Analyses were conducted for both unadjusted and adjusted for cotinine levels (in the case of associations of genotype with daily cigarette consumption) and cigarette consumption (in the case of associations of genotype with cotinine levels). In the Midspan Family Study, which is one of the Midspan studies (see Supplementary Methods, available online), within-family clustering was adjusted for by including a family-level random intercept (42). We assumed an additive model of genetic action, and a linear relationship between cigarette consumption and cotinine level, consistent with previous reports (3,43). The rs1051730 and rs16969968 variants are in perfect linkage disequilibrium in HapMap3 samples of European ancestry ($D' = 1.0$, $R^2 =$ 1.0) (http://hapmap.ncbi.nlm.nih.gov/), and therefore, these interchangeable SNPs were considered as a single marker in these analyses. These per-allele associations were pooled using a random effects method (44). Random effects models are typically more conservative than fixed-effects models, although in the absence of substantial between-study heterogeneity, the two methods generate similar results and generate identical results where there is perfect homogeneity. The I^2 statistic was used to estimate the percentage of total variation in study estimates resulting from between-study heterogeneity. Conventionally, I^2 values less than 25% are considered low and unlikely to represent important heterogeneity (45). The Cochran Q test was used to evaluate the statistical evidence for between-study heterogeneity.

To assess the congruence of our results with odds ratios (ORs) for lung cancer reported for rs1051730–rs16969968 genotype (1), data were used from a large case–control study that assessed the relationship between serum cotinine levels and lung cancer (46). This study reported odds ratios and 95% confidence intervals (CIs) for eight separate cotinine intervals with reference to the lowest cotinine range, produced using conditional logistic regression with adjustment for sex, year of birth, time of enrollment, and geographical region, and with adjustment for measurement error using repeated cotinine measurements on a subset of samples. These eight odds ratios and confidence intervals were abstracted, together with the midpoint of each cotinine interval, converted to nmol/L, with the midpoint of the highest interval taken as 400 ng/mL as in the original report. Odds ratios and their confidence intervals were logarithmically transformed so that log odds ratios and their standard errors were calculated. The association between cotinine levels and lung cancer risk is shown in Supplementary Figure 1 (available online). Weighted least squares regression was then carried out on log odds ratios, weighted by the

inverse of the standard error squared, with the intercept forced to be 0 to represent the odds ratio of 1 for the lowest cotinine interval. The resulting regression coefficient was then multiplied by the per-allele increase in cotinine estimated from this study and converted to an expected odds ratio for lung cancer risk associated with each additional copy of the rs1051730–rs16969968 risk allele. Lower and upper limits of the confidence interval from the per-allele increase were also calculated in the same way.

All analyses were conducted with Stata software (version 11.2; StataCorp, College Station TX). All statistical tests were two-sided, except where stated, and all P values of .05 or less were considered statistically significant.

Results

Characteristics of Participants

A total of 12 364 participants of self-reported European ancestry, with complete smoking status (including cigarette consumption among current smokers), cotinine levels, and genotype data, were available across six studies (see Table 1). Participants were classified as never smokers (n = 4771), former smokers (n = 4661), and current smokers (n = 2932) based on their self-reported smoking behavior (see Table 2).

Smoking Status, Genotype, and Cotinine Level

Cotinine levels confirmed the classification of participants as nonsmokers (ie, never smokers and former smokers) and current smokers (Table 2), based on conventional indicative cotinine levels among nonsmokers and current smokers (never smokers and former smokers, approximately 5 nmol/L; current smokers, >85 nmol/L) (47). Across all six studies, median cotinine levels for nonsmokers were 3.40 nmol/L or lower (ranging from 0.57 to 3.40 nmol/L), whereas median cotinine levels in current smokers were 968.5 nmol/L or higher (ranging from 968.5 to 1535.0 nmol/L). As described previously, cotinine levels in current smokers in the EPIC study differed between case subjects and control subjects and were therefore considered separately.

Our analysis was based on current smokers to ascertain the relationship between rs1051730–rs16969968 genotype and both daily cigarette consumption and cotinine levels. Median cotinine levels in current smokers increased with each copy of the minor allele (rs1051730 T, rs16969968 A), typically by about 140 nmol/L (range = 45–255 nmol/L) (Table 2). Individual study associations between genotype and both daily cigarette consumption and cotinine levels are reported in Supplementary Table 1 (available online).

Association of rs1051730–rs16969968 Genotype With Daily Cigarette Consumption

Meta-analysis of the per-allele association of rs1051730–rs16969968 genotype on unadjusted self-reported daily cigarette consumption among current smokers indicated strong evidence of association (Figure 1). Pooling the six studies within a random effects framework indicated that the risk allele was associated with increased self-reported cigarette consumption (mean increase in number of cigarettes per day per allele = 1.0 cigarette, 95% CI = 0.57 to 1.43 cigarettes, $P = 5.22 \times 10^{-6}$). The between-study heterogeneity was low and not statistically significant ($I^2 = 0\%$, $P_{heterogeneity} = .72$).

Table 1. Characteristics of participants by study*

	Study						
Characteristic	**BRHS** (n = 3613)	**BWHHS** (n = 3684)	**EPIC Case†** (n = 758)	**EPIC Control†** (n = 1548)	**Midspan** (n = 1872)	**Patch II** (n = 451)	**PiP (n = 438)**
Sex, No. (%)							
Male	3613 (100%)	—	687 (64%)	948 (61%)	855 (46%)	166 (37%)	221 (51%)
Female	—	3684 (100%)	71 (36%)	600 (39%)	1017 (54%)	285 (63%)	217 (49%)
Age, mean (range), y	69 (58–81)	69 (59–80)	58 (34–78)	58 (35–79)	45 (30–59)	51 (33–73)	44 (19–78)
Genotype‡	rs1051730	rs1051730	rs1699698	rs1699698	rs1051730	rs1051730	rs1051730
Biological material	Serum	Serum	Serum	Serum	Serum	Plasma	Plasma

* Data on genotype (rs16969968 or rs1051730), smoking status (never smoker, former smoker, and current smoker), and cotinine levels determined in blood (plasma or serum), as well as cigarette consumption (cigarette per day) among current smokers, in participants of European ancestry were available from six independent studies (BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; EPIC = European Prospective Investigation into Cancer and Nutrition; PiP = Patch in Practice).

† Cotinine levels and self-reported cigarette consumption differed in current smoker case subjects (EPIC Case) and control subjects (EPIC Control). Mean cotinine levels were higher in the EPIC Case subsample (1391 nmol/L, SD = 588) than the EPIC Control subsample (1006 nmol/L, SD = 589) even after adjustment for cigarette consumption (Kruskal–Wallis $P < .001$, two-sided), indicating possible differences in smoking behavior (eg, depth of inhalation). This difference at baseline may have contributed to the likelihood of a subsequent lung cancer diagnosis. Therefore, the EPIC case subjects and control subjects were analyzed and presented separately in this study.

‡ Genotype: rs1051730 (wild-type allele C, risk allele T); rs16969968 (wild-type allele G, risk allele A).

After adjustment for cotinine levels, the per-allele estimate was reduced by 50% (mean increase in number of cigarettes per day per allele = 0.45, 95% CI = 0.05 to 0.86 cigarettes, $P = .029$). Between-study heterogeneity remained low and was not statistically significant ($I^2 = 0\%$, $P_{heterogeneity} = .51$). Weak evidence for a difference between the estimates for the unadjusted and adjusted analyses was observed ($P = .07$) (not shown in the figure).

Association of rs1051730-rs16969968 With Cotinine Levels

Meta-analysis of the per-allele association of rs1051730–rs16969968 genotype on unadjusted cotinine levels among current smokers indicated strong evidence of association (Figure 2). Pooling the six studies within a random effects framework indicated that the risk allele was associated with increased cotinine levels (mean increase in cotinine levels per allele = 138.72 nmol/L, 95% CI = 97.91 to 179.53 nmol/L, $P = 2.71 \times 10^{-11}$). The between-study heterogeneity was low and not statistically significant ($I^2 = 23\%$, $P_{heterogeneity} = .26$). Following adjustment for self-reported cigarette consumption, the per-allele estimate was reduced by only 18% (mean increase in cotinine levels per allele = 113.76 nmol/L, 95% CI = 76.88 to 150.64 nmol/L, $P = 1.49 \times 10^{-9}$). Between-study heterogeneity remained low and was not statistically significant ($I^2 = 16\%$, $P_{heterogeneity} = .31$). No evidence for a difference between the estimates for the unadjusted and adjusted analyses was observed ($P = .37$) (not shown in the figure).

Association of rs1051730–rs16969968 Genotype With Lung Cancer Risk

To estimate the extent to which the association of rs1051730–rs16969968 genotype with lung cancer is mediated via smoking, we applied the effect size we observed for the association with cotinine levels to published data on the association between cotinine levels and lung cancer risk. Applying weighted least squares regression to the log odds ratios presented by Boffetta et al. (46) produced a regression coefficient of 0.001945 per nmol/L of serum cotinine for the association between cotinine levels and lung cancer risk

(Supplementary Figure 1, available online). Thus, the observed per-allele increase in cotinine levels of 138.72 nmol/L (95% CI = 97.91 to 179.53) would indicate an increased risk of lung cancer with each additional copy of the rs1051730–rs16969968 risk allele (per-allele OR = 1.31, 95% CI = 1.21 to 1.42). This corresponds closely with published data for the association between rs1051730–rs16969968 genotype and lung cancer risk, where an odds ratio of 1.32 has been reported (1).

Sample Size Calculation for GWAS

We next assessed the implications of the observed stronger association of rs1051730–rs16969968 genotype with cotinine levels, compared with self-reported daily cigarette consumption, for the design of future GWAS. The association of rs1051730–rs16969968 with self-reported daily cigarette consumption and cotinine levels indicated by our results suggests that a GWAS would require a sample size in excess of 7000 to detect an association of the rs1051730–rs16969968 variant with self-reported daily cigarette consumption, but a sample size of only 1800–1900 to detect an association with cotinine levels, assuming an additive model for the genetic association, a minor allele frequency of 0.35, an alpha level of 5×10^{-8}, and a required power of at least 80%. In addition, a recall-by-genotype design, where rs1051730–rs16969968 homozygotes are preselected, would require a sample size of only 200 to detect this association with cotinine levels, with an alpha level of 0.05, and a required power of at least 80%.

Discussion

Our data show that rs1051730–rs16969968 genotype is strongly associated with tobacco exposure measured objectively via cotinine levels and that this association is robust even after adjustment for self-reported cigarette consumption. We used the per-allele association of genotype on cotinine levels to estimate the association between rs1051730–rs16969968 genotype and lung cancer risk, using published data on the association between cotinine levels and

Table 2. Cotinine levels by smoking behavior and genotype in current smokers*
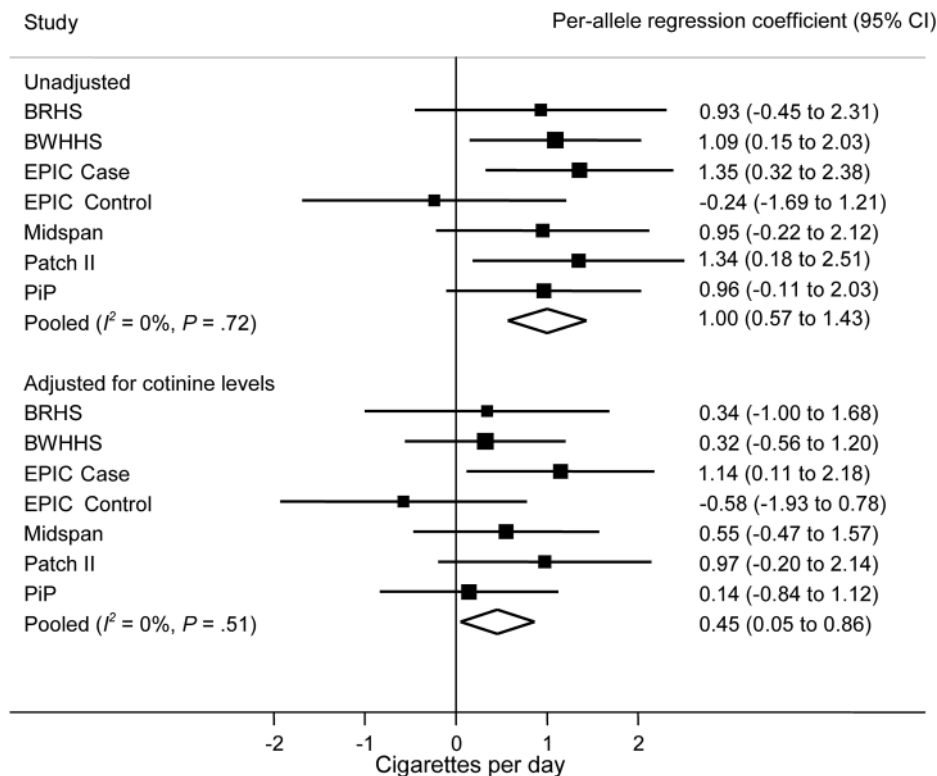
Smoking behavior or No. of risk alleles in current smokers†	BRHS Participants, No. (%)	BRHS Cotinine, median (IQR), nmol/L	BWHHS Participants, No. (%)	BWHHS Cotinine, median (IQR), nmol/L	EPIC Case‡ Participants, No. (%)	EPIC Case‡ Cotinine, median (IQR), nmol/L	EPIC Control‡ Participants, No. (%)	EPIC Control‡ Cotinine, median (IQR), nmol/L	Midspan Participants, No. (%)	Midspan Cotinine, median (IQR), nmol/L	Patch II Participants, No. (%)	Patch II Cotinine, median (IQR), nmol/L	PiP Participants, No. (%)	PiP Cotinine, median (IQR), nmol/L
Smoking status														
Never Smokers	1066 (30%)	0.79 (0.28–2.16)	2081 (56%)	0.57 (0.28–1.31)	92 (12%)	1.58 (0.21–3.00)	685 (44%)	1.76 (0.60–4.07)	847 (45%)	2.84 (1.70–5.67)	0 (0%)		0 (0%)	
Former Smokers	2162 (60%)	1.13 (0.28–4.31)	1203 (33%)	0.80 (0.28–2.22)	221 (29%)	2.25 (0.92–5.01)	549 (35%)	1.88 (0.82–4.49)	526 (28%)	3.40 (1.70–9.08)	0 (0%)		0 (0%)	
Current Smokers	385 (10%)	1330.2 (791.1–1814.2)	400 (11%)	1219.5 (745.8–1786.4)	445 (59%)	1401.5 (1036.4–1767.7)	314 (20%)	968.5 (507.5–1386.8)	499 (27%)	1444.2 (800.2–1896.0)	451 (100%)	1535.0 (1167.6–1945.9)	438 (100%)	1526.2 (1102.2–2042.4)
Number of risk alleles														
0	175 (45%)	1171.8 (675.3–1735.4)	171 (43%)	1090 (645.8–1612.6)	151 (34%)	1272.2 (936.2–1644.8)	129 (41%)	930.17 (491.3–1310)	242 (48%)	1383.8 (764.5–1809.3)	190 (42%)	1464 (1090.3–1879.1)	191 (44%)	1421.7 (1044.6–1944.3)
1	171 (45%)	1385.8 (818.3–1956.1)	185 (46%)	1264.7 (790.7–1902.2)	203 (46%)	1498.8 (1059.1–1804.1)	145 (46%)	991.8 (548.7–1386.8)	201 (41%)	1484.5 (840.4–1942.5)	198 (44%)	1539.9 (1207.4–1904.3)	198 (45%)	1559.4 (1125.8–2075.3)
2	39 (10%)	1681.4 (1243.9–2072.4)	44 (11%)	1399.6 (964.5–2027.2)	91 (20%)	1504.2 (1139.9–1843.1)	40 (13%)	1021.4 (556.6–1629.9)	56 (11%)	1522.6 (832.2–1971.0)	63 (14%)	1787.7 (1338.9–2151.2)	49 (11%)	1847.1 (1381.4–2264.6)

* BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; EPIC = European Prospective Investigation into Cancer and Nutrition; IQR = interquartile range between 25th and 75th percentile; PiP = Patch in Practice.

† Typical cotinine values (never smokers and former smokers approximately 5 nmol/L; current smokers >85 nmol/L) [47]. Number of risk alleles expressed as zero, one, or two copies of the rs1051730 T allele or rs16969968 A allele.

‡ Cotinine levels and self-reported cigarette consumption differed in current smoker case subjects (EPIC Case) and control subjects (EPIC Control). Cotinine levels were higher in the EPIC Case subsample (mean level = 1391 nmol/L, SD = 588) than the EPIC Control subsample (mean level = 1006 nmol/L, SD = 589) even after adjustment for cigarette consumption (two-sided Kruskal–Wallis P < .001), indicating possible differences in smoking behavior (eg, depth of inhalation). This difference at baseline may have contributed to the likelihood of a subsequent lung cancer diagnosis. Therefore, the EPIC case subjects and control subjects were analyzed and presented separately in this study.

Figure 1. Meta-analysis of association of rs1051730–rs16969968 risk allele with cigarette consumption in current smokers. Data from six independent studies contributed to the meta-analysis. Cotinine levels and self-reported cigarette consumption differed in current smoker case subjects (EPIC Case) and control subjects (EPIC Control) in the EPIC study, so they were analyzed separately. In each study, linear regression was used to calculate per-allele association of rs1051730–rs16969968 genotype with daily cigarette consumption. Units represent cigarettes per day. Unadjusted and adjusted (for cotinine levels) analyses are shown. The I^2 statistic was used to estimate the percentage of total variation in study estimates resulting from between-study heterogeneity. Individual study regression coefficients were combined using random effects methods. **Squares** represent per-allele regression coefficients, which represent mean increase in number of cigarettes per day per allele; **size of the square** represents inverse of the variance of the regression coefficient; **horizontal lines** represent 95% CIs; **diamonds** represent summary estimate combining the study-specific estimates using a random effects model; **solid vertical line** represents a regression coefficient of 0. P for heterogeneity was derived from the Cochran Q test (one-sided). All other statistical tests were two-sided, and statistical significance required a P value of .05 or less. BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; CI = confidence interval; EPIC = European Prospective Investigation into Cancer and Nutrition; PiP = Patch in Practice.

Study	Per-allele regression coefficient (95% CI)
Unadjusted	
BRHS	0.93 (−0.45 to 2.31)
BWHHS	1.09 (0.15 to 2.03)
EPIC Case	1.35 (0.32 to 2.38)
EPIC Control	−0.24 (−1.69 to 1.21)
Midspan	0.95 (−0.22 to 2.12)
Patch II	1.34 (0.18 to 2.51)
PiP	0.96 (−0.11 to 2.03)
Pooled (I^2 = 0%, P = .72)	1.00 (0.57 to 1.43)
Adjusted for cotinine levels	
BRHS	0.34 (−1.00 to 1.68)
BWHHS	0.32 (−0.56 to 1.20)
EPIC Case	1.14 (0.11 to 2.18)
EPIC Control	−0.58 (−1.93 to 0.78)
Midspan	0.55 (−0.47 to 1.57)
Patch II	0.97 (−0.20 to 2.14)
PiP	0.14 (−0.84 to 1.12)
Pooled (I^2 = 0%, P = .51)	0.45 (0.05 to 0.86)

lung cancer risk. Our estimate of the association between genotype and lung cancer risk was consistent with previously reported estimates, even though we were only able to capture point prevalence smoking intensity (ie, based on current smoking only), and not lifetime exposure. These data therefore support the conclusion that association of rs1051730–rs16969968 genotype with lung cancer risk is mediated largely, if not wholly, via tobacco exposure. Although some studies have suggested a direct contribution of rs1051730–rs16969968 genotype to lung carcinogenesis, these have typically relied on self-report measures of smoking behavior which, as we have shown, do not fully capture actual exposure.

The association between these variants and lung cancer constitutes Mendelian randomization evidence on the causal nature of the smoking–lung cancer association (48). In this case, we do not, of course, require such confirmation, but it serves as proof of principle for this approach. These findings also have important implications for epidemiology and genetic association studies, including large GWAS of cigarette smoking behavior, which typically rely on retrospective self-report measures. We discuss these implications below.
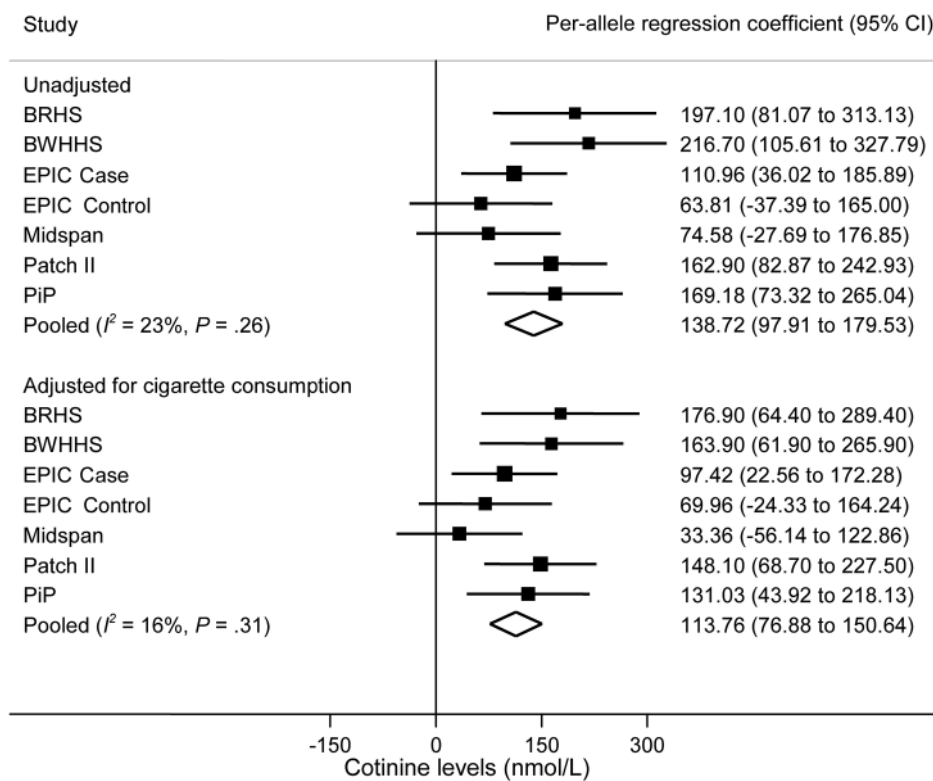
It is now well established that smokers modify their smoking behavior to self-titrate circulating nicotine to a level appropriate to their need (49). This compensatory behavior is achieved through varying the number of puffs, puff volume and interpuff interval, and covering the cigarette filter to reduce ventilation by sidestream air. This plasticity of smoking behavior means that estimating exposure to nicotine and tar in cigarette smokers is not possible

through the use of machine protocols to calculate yield estimates (50) or through the simple counting of number and strength of cigarettes smoked. Smokers are able to titrate not only how many cigarettes they smoke but what strength of cigarette they smoke and how they smoke them. Therefore, either biochemical measures of exposure or naturalistic measures of smoking topography are necessary if an acceptable level of measurement precision is to be achieved (51).

By extension, GWAS that rely on self-report measures to quantify smoking behavior, and therefore tobacco exposure, may be insensitive to relatively modest genetic associations. Recent studies of smoking phenotypes have enjoyed considerable success in identifying loci associated with various aspects of smoking behavior (9,20,22). Our results suggest that these studies may have underestimated the magnitude of these associations (perhaps, in particular, in the case of heaviness of smoking phenotypes). It is also likely that a number of common variants that contribute an important proportion of phenotypic variance may remain unidentified. Using biomarker phenotypes, such as cotinine, to assay nicotine consumption and tobacco exposure may improve the success of GWAS and identify additional novel variants associated with nicotine consumption.

Our power analysis also indicates that these results have important implications for the conduct of GWAS with respect to sample size. Considerably, smaller sample sizes may be sufficient to detect robust genome-wide associations if these use better outcome measures—there is a trade-off between sample size and phenotype

Figure 2. Meta-analysis of association of rs1051730–rs16969968 risk allele with cotinine levels in current smokers. Data from six independent studies contributed to the meta-analysis. Cotinine levels and self-reported cigarette consumption differed in current smoker case subjects (EPIC Case) and control subjects (EPIC Control) in the EPIC study, so they were analyzed separately. In each study, linear regression was used to calculate per-allele association of rs1051730–rs16969968 genotype with cotinine levels. Units represent nmol/L. Unadjusted and adjusted (for cigarette consumption) analyses are shown. The I^2 statistic was used to estimate the percentage of total variation in study estimates resulting from between-study heterogeneity. Individual study regression coefficients were combined using random effects methods. **Squares** represent per allele regression coefficient, which represents mean increase in cotinine levels per allele; **size of the square** represents inverse of the variance of the regression coefficient; **horizontal lines** represent 95% CIs; **diamonds** represent summary estimate combining the study-specific estimates with a random effects model; **solid vertical line** represents a regression coefficient of 0. *P* for heterogeneity was derived from the Cochran *Q* test (one-sided). All other statistical tests were two-sided, and statistical significance required a *P* value of .05 or less. BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; CI = confidence interval; EPIC = European Prospective Investigation into Cancer and Nutrition; PiP = Patch in Practice.

Study	Per-allele regression coefficient (95% CI)
Unadjusted	
BRHS	197.10 (81.07 to 313.13)
BWHHS	216.70 (105.61 to 327.79)
EPIC Case	110.96 (36.02 to 185.89)
EPIC Control	63.81 (-37.39 to 165.00)
Midspan	74.58 (-27.69 to 176.85)
Patch II	162.90 (82.87 to 242.93)
PiP	169.18 (73.32 to 265.04)
Pooled (I^2 = 23%, *P* = .26)	138.72 (97.91 to 179.53)
Adjusted for cigarette consumption	
BRHS	176.90 (64.40 to 289.40)
BWHHS	163.90 (61.90 to 265.90)
EPIC Case	97.42 (22.56 to 172.28)
EPIC Control	69.96 (-24.33 to 164.24)
Midspan	33.36 (-56.14 to 122.86)
Patch II	148.10 (68.70 to 227.50)
PiP	131.03 (43.92 to 218.13)
Pooled (I^2 = 16%, *P* = .31)	113.76 (76.88 to 150.64)

Cotinine levels (nmol/L): -150, 0, 150, 300

quality and/or precision. It may not always be practical or financially possible to collect such phenotypes in large samples. An initial GWAS based on a preliminary phenotype, which is easy to collect (such as cigarette consumption), can be followed up by high-quality/precision phenotyping (such as cotinine measurements) in a sample selected by genotype on which the stored biological specimens are available. This combination could lead to considerable increases in statistical power and efficiency. In conventional observational epidemiology, where associations between self-reported smoking behavior and outcomes underestimate the actual etiological associations that exist (52), more detailed phenotyping reveals stronger associations (46). Residual associations between other exposures and smoking-related outcomes, such as lung cancer, which persist after statistical control for self-reported cigarette consumption, should therefore be treated with caution (53).

There are some limitations to consider when interpreting these results. First, our data are drawn from disparate studies recruited from various populations. Nevertheless, the consistency in effect size estimates across samples, and the lack of substantial between-study heterogeneity, suggests that the impact of this is minimal. Second, our estimate of the strength of the association between genotype and cotinine exposure on lung cancer risk relies on an indirect comparison with published data. Moreover, it relies on measures of current smoking rather than lifetime exposure, which is more strongly associated with lung cancer risk. Nevertheless, the triangulation of our data with published estimates of the association

of genotype with lung cancer risk raises confidence in these results. Third, we lacked repeated measurement of cotinine, which would allow the assessment of within- and between-person variation (eg, due time between last cigarette smoked and collection of biological sample for cotinine analysis), and the time of day that cotinine samples were collected was not standardized, either within or between samples. However, the relatively long half-life of cotinine means it provides a reasonably stable measure of exposure in regular cigarette smokers. Also, the lack of substantial between-study heterogeneity again suggests that the impact of this was modest. Fourth, we were unable to capture interindividual variation in cotinine metabolism, for example due to genetic variation within cytochrome P450, family 2, subfamily A, polypeptide 6 (*CYP2A6*), which encodes the CYP2A6 enzyme primarily responsible for the metabolism of nicotine to cotinine (54). Therefore, although cotinine levels provide a considerably more precise measure of tobacco exposure than self-report measures, including measures of CYP2A6 activity might serve to refine this further. It is also possible that increased nicotine consumption among rs1051730–rs16969968 risk allele carriers may result in CYP2A6 enzyme induction, giving rise to a positive feedback cycle as faster metabolism of nicotine leads to increased consumption to maintain circulating levels (55). However, evidence suggests that nicotine metabolism is inhibited in smokers compared with nonsmokers (56). We also did not adjust for other factors known to influence nicotine metabolism, such as sex and body mass index.

In conclusion, our data indicate that much of the debate regarding whether the association of rs1051730–rs16969968 genotype with lung cancer is direct or operate indirectly via tobacco exposure is essentially due to imprecision in the measures of tobacco use and exposure used in most studies. More importantly, our results show that the search for even larger sample sizes in GWAS may generate diminishing returns if this is at the expense of phenotype precision. The use of objective measures of smoking behavior in genome-wide studies may reveal novel variants associated with these outcomes, which would be undetectable using conventional self-report measures.

References

1. Amos CI, Wu XF, 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):638–639.
4. Pillai SG, Ge DL, Zhu GH, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009;5(3):e1000421.
5. Young RP, Hophins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *Eur Resp J.* 2008;32(5):1158–1164.
6. Zintzaras E, Zdoukopoulos N. A field synopsis and meta-analysis of genetic association studies in peripheral arterial disease: the CUMAGAS-PAD database. *Am J Epidemiol.* 2009;170(1):1–11.
7. Freathy RM, Ring SM, Shields B, et al. A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet.* 2009;18(15):2922–2927.
8. Freathy RM, Kazeem GR, Morris RW, et al. Genetic variation at CHRNA5-CHRNA3-CHRNB4 interacts with smoking status to influence body mass index. *Int J Epidemiol.* 2011;40(6):1617–1628.
9. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010; 42(5):436–440.
10. Shiraishi K, Kohno T, Kunitoh H, et al. Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. *Carcinogenesis.* 2009;30(1):65–70.
11. Baker TB, Weiss RB, Bolt D, et al. Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine Tob Res.* 2009;11(7):785–796.
12. Breitling LP, Dahmen N, Mittelstrass K, et al. Association of nicotinic acetylcholine receptor subunit alpha 4 polymorphisms with nicotine dependence in 5500 Germans. *Pharmacogenomics J.* 2009;9(4):219–224.
13. Broderick P, Wang YF, Vijayakrishnan J, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res.* 2009;69(16):6633–6641.
14. Chen XN, Chen JC, Williamson VS, et al. Variants in nicotinic acetylcholine receptors alpha 5 and alpha 3 increase risks to nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet.* 2009;150B(7):926–933.
15. Lips EH, Gaborieau V, McKay JD, et al. Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol.* 2010;39(2):563–577.
16. Saccone NL, Wang JC, Breslau N, et al. The CHRNA5-CHRNA3-CHRNB4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.* 2009;69(17):6848–6856.
17. Schlaelpfer IR, Hoft NR, Collins AC, et al. The CHRNA5/A3/B4 gene cluster variability as an important determinant of early alcohol and tobacco initiation in young adults. *Biol Psychiatry.* 2008;63(11):1039–1046.
18. Sherva R, Wilhelmsen K, Pomerleau CS, et al. Association of a single nucleotide polymorphism in neuronal acetylcholine receptor subunit alpha 5 (CHRNA5) with smoking status and with 'pleasurable buzz' during early experimentation with smoking. *Addiction.* 2008;103(9):1544–1552.
19. Stevens VL, Bierut LJ, Talbot JT, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev.* 2008;17(12):3517–3525.
20. Tobacco-and-Genetics-Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5): 441–447.
21. Ware JJ, van den Bree MB, Munafo MR. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. *Nicotine Tob Res.* 2011;13(12):1167–1175.
22. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet.* 2010;42(5):448–453.
23. 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.
24. Berrettini W, Yuan X, Tozzi F, et al. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry.* 2008;13(4): 368–373.
25. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry.* 2008;165(9):1163–1171.
26. Shipton D, Tappin DM, Vadiveloo T, Crossley JA, Aitken DA, Chalmers J. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. *BMJ.* 2009;339:b4347.
27. Keskitalo K, Broms U, Heliovaara M, et al. Association of serum cotinine level with a cluster of three nicotinic acetylcholine receptor genes (CHRNA3/CHRNA5/CHRNB4) on chromosome 15. *Hum Mol Genet.* 2009;18(20):4007–4012.
28. Le Marchand L, Derby KS, Murphy SE, et al. Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res.* 2008;68(22):9137–9140.
29. Aveyard P, Brown K, Saunders C, et al. Weekly versus basic smoking cessation support in primary care: a randomised controlled trial. *Thorax.* 2007;62(10):898–903.
30. Baltar VT, Xun WW, Chuang SC, et al. Smoking, second-hand smoke and cotinine levels in a subset of EPIC cohort. *Cancer Epidemiol Biomarkers Prev.* 2011;20(5):869–875.
31. Hart CL, MacKinnon PL, Watt GC, et al. The Midspan studies. *Int J Epidemiol.* 2005;34(1):28–34.
32. Imperial-Cancer-Research-Fund. Effectiveness of a nicotine patch in helping people stop smoking: results of a randomised trial in general practice. *BMJ.* 1993;306(6888):1304–1308.
33. Johansson M, Relton C, Ueland PM, et al. Serum B vitamin levels and risk of lung cancer. *JAMA.* 2010;303(23):2377–2385.
34. Johnstone E, Benowitz N, Cargill A, et al. Determinants of the rate of nicotine metabolism and effects on smoking behavior. *Clin Pharmacol Ther.* 2006;80(4):319–330.
35. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *J Epidemiol Community Health.* 2003;57(2):134–140.
36. Middtun O, Hustad S, Ueland PM. Quantitative profiling of biomarkers related to B-vitamin status, tryptophan metabolism and inflammation in human plasma by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2009;23(9):1371–1379.
37. Munafo MR, Johnstone E, Walther D, Uhl GR, Murphy M, Aveyard P. CHRNA3 rs1051730 genotype and short-term smoking cessation. *Nicotine Tob Res.* 2011;13(10):982–988.

38. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* 2002;5(6B):1113–1124.

39. Walker M, Whincup PH, Shaper AG. The British Regional Heart Study 1975–2004. *Int J Epidemiol.* 2004;33(6):1185–1192.

40. Yudkin P, Hey K, Roberts S, Welch S, Murphy M, Walton R. Abstinence from smoking eight years after participation in randomised controlled trial of nicotine patch. *BMJ.* 2003;327(7405):28–29.

41. Timofeeva MN, McKay JD, Smith GD, et al. Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiol Biomarkers Prev.* 2011;20(10):2250–2261.

42. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS.* New York, NY: Springer; 2000.

43. Etter JF, Vu Duc T, Perneger TV. Saliva cotinine levels in smokers and nonsmokers. *Am J Epidemiol.* 2000;151(3):251–258.

44. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7(3):177–188.

45. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557–560.

46. Boffetta P, Clark S, Shen M, Gislefoss R, Peto R, Andersen A. Serum cotinine level as predictor of lung cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2006;15(6):1184–1188.

47. Society-for-Research-on-Nicotine-and-Tobacco. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res.* 2002;4(2):149–159.

48. Davey-Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.

49. Strasser AA, Lerman C, Sanborn PM, Pickworth WB, Feldman EA. New lower nicotine cigarettes can produce compensatory smoking and increased carbon monoxide exposure. *Drug Alcohol Depend.* 2007;86(2–3):294–300.

50. Parish S, Collins R, Peto R, et al. Cigarette smoking, tar yields, and non-fatal myocardial infarction: 14,000 cases and 32,000 controls in the United Kingdom. The International Studies of Infarct Survival (ISIS) Collaborators. *BMJ.* 1995;311(7003):471–477.

51. Hammond D, Fong GT, Cummings KM, O'Connor RJ, Giovino GA, McNeill A. Cigarette yields and human exposure: a comparison of alternative testing regimens. *Cancer Epidemiol Biomarkers Prev.* 2006;15(8):1495–1501.

52. Hart CL, Davey-Smith G, Hole DJ, Hawthorne VM. Carboxyhaemoglobin concentration, smoking habit, and mortality in 25 years in the Renfrew/Paisley prospective cohort study. *Heart.* 2006;92(3):321–324.

53. Phillips AN, Davey-Smith G. Bias in relative odds estimation owing to imprecise measurement of correlated exposures. *Stat Med.* 1992;11(7):953–961.

54. Tutka P, Mosiewicz J, Wielosz M. Pharmacokinetics and metabolism of nicotine. *Pharmacol Rep.* 2005;57(2):143–153.

55. Tyndale RF, Sellers EM. Variable CYP2A6-mediated nicotine metabolism alters smoking behavior and risk. *Drug Metab Dispos.* 2001;29(4, pt 2):548–552.

56. Mwenifumbo JC, Sellers EM, Tyndale RF. Nicotine metabolism and CYP2A6 activity in a population of black African descent: impact of gender and light smoking. *Drug Alcohol Depend.* 2007;89(1):24–33.

Affiliations of authors: School of Experimental Psychology, University of Bristol, Bristol, UK (MRM); Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon CEDEX 08, France (MNT, PB); Division of Population Health, University College London, London, UK (RWM, JPC, BJJ, SE); Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK (DP-M, JPC, AA); British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, UK (NS, SP); Department of Oncology, University of Oxford, Oxford, UK (ECJ); Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK (CR); Robertson Centre for Biostatistics, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK (PCDJ); Molecular Neurobiology Branch, NIH-IRP (NIDA), Baltimore, USA (DW, GRU); Division of Population Health Sciences and Education, St George's University of London, London, UK (PHW); School of Public Health, Imperial College, London, UK (PV, ER); HuGeF Foundation, Turin, Italy (PV); Woodlands Family Medical Centre, Stockton on Tees, UK (MNU); Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK (PA); University College London, London, UK (ADH); Division of Community-Based Sciences, University of Glasgow, Glasgow, UK (GW); MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE), School of Social and Community Medicine, University of Bristol, Bristol, UK (TMP, NJT, GDS).