Munaf, MR; Timofeeva, MN; Morris, RW; Prieto-Merino, D; Sattar, N; Brennan, P; Johnstone, EC; Relton, C; Johnson, PC; Walther, D; Whincup, PH; Casas, JP; Uhl, GR; Vineis, P; Padmanabhan, S; Jefferis, BJ; Amuzu, A; Riboli, E; Upton, MN; Aveyard, P; Ebrahim, S; Hingorani, AD; Watt, G; Palmer, TM; Timpson, NJ; EPIC Study Group, ; Davey Smith, G (2012) Association between genetic variants on chromosome 15q25 locus and objective measures of tobacco exposure. J Natl Cancer Inst, 104 (10). pp. 740-8. ISSN 0027-8874

Downloaded from: http://researchonline.lshtm.ac.uk/20890/

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


Manuscript received September 12, 2011; revised February 29, 2012; accepted March 12, 2012.

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 12364 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, \( P = 5.22 \times 10^{-2} \)) 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...
SNPs should be more strongly underappreciated. 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), 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 cotinine 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 (\( D^2 = 1.0, R^2 = 1.0 \) (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 \( F \) statistic was used to estimate the percentage of total variation in study estimates resulting from between-study heterogeneity. Conventionally, \( F \) 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 96.5 nmol/L or higher (ranging from 96.8 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, \( P = 5.22 \times 10^{-8} \)). The between-study heterogeneity was low and not statistically significant (\( I^2 = 0\% \), \( P_{\text{heterogeneity}} = .72 \)).
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 (\( F = 0 \), \( P_{\text{heterogeneity}} = .51 \)). 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 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 (\( F = 23 \), \( P_{\text{heterogeneity}} = .26 \)). 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 (\( F = 16 \), \( P_{\text{heterogeneity}} = .31 \)). 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 \( \times 10^{-4} \), 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

---

* 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).
Table 2. Cotinine levels by smoking behavior and genotype in current smokers*  

<table>
<thead>
<tr>
<th>Smoking behavior or No. of risk alleles in current smokers†</th>
<th>BRHS</th>
<th>BWHHS</th>
<th>EPIC Case‡</th>
<th>EPIC Control‡</th>
<th>Midspan</th>
<th>Patch II</th>
<th>PiP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1066 (30%)</td>
<td>0.79 (0.28-2.16)</td>
<td>2081 (56%)</td>
<td>0.57 (0.28-1.31)</td>
<td>92 (12%)</td>
<td>1.58 (0.21-3.00)</td>
<td>685 (44%)</td>
</tr>
<tr>
<td>Former</td>
<td>2162 (60%)</td>
<td>1.13 (0.28-4.31)</td>
<td>1203 (33%)</td>
<td>0.80 (0.28-2.23)</td>
<td>221 (29%)</td>
<td>2.25 (0.92-5.01)</td>
<td>549 (35%)</td>
</tr>
<tr>
<td>Current</td>
<td>385 (10%)</td>
<td>1.33 (0.791-1.814)</td>
<td>400 (11%)</td>
<td>1.219 (745.8-1786.4)</td>
<td>445 (59%)</td>
<td>1.401 (1036.4-1767.7)</td>
<td>314 (20%)</td>
</tr>
<tr>
<td>Number of risk alleles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>177 (45%)</td>
<td>171 (117.1-171.8)</td>
<td>171 (43%)</td>
<td>1090 (645.8-1612.6)</td>
<td>151 (34%)</td>
<td>1272.2 (936.2-1648.4)</td>
<td>129 (41%)</td>
</tr>
<tr>
<td>1</td>
<td>171 (45%)</td>
<td>1385.6 (818.3-1956.1)</td>
<td>185 (46%)</td>
<td>1264.7 (790.7-1902.2)</td>
<td>203 (46%)</td>
<td>1498.8 (1059.1-1804.1)</td>
<td>145 (46%)</td>
</tr>
<tr>
<td>2</td>
<td>39 (10%)</td>
<td>1681.4 (1243.9-2072.6)</td>
<td>44 (11%)</td>
<td>1399.6 (964.5-2027.2)</td>
<td>91 (20%)</td>
<td>1504.2 (1139.9-1843.1)</td>
<td>40 (13%)</td>
</tr>
</tbody>
</table>

* 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.
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 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
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 \((\text{CYP2A6})\), 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.

**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 \(P\) 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.
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


Funding

Wellcome Trust project (086684 to MRM) and a postdoctoral International Agency for Research on Cancer fellowship (to MNT).