UCSF
UC San Francisco Previously Published Works

Title

Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls

Permalink

https://escholarship.org/uc/item/4b10x00r

Journal

Human Genetics, 132(5)

ISSN

0340-6717

Authors

Brenner, Darren R
Brennan, Paul
Boffetta, Paolo
et al.

Publication Date

2013-05-01

DOI

10.1007/s00439-013-1270-y

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

NIH Public Access
Author Manuscript
Hum Genet. Author manuscript; available in PMC 2014 May 01.

Published in final edited form as:
Hum Genet. 2013 May ; 132(5): 579–589. doi:10.1007/s00439-013-1270-y.

Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls

Darren R. Brenner,
International Agency for Research on Cancer, Lyon, France

Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON, Canada

Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Paul Brennan,
International Agency for Research on Cancer, Lyon, France

Paolo Boffetta,
Tisch Cancer Institute, Mount Sinai School of Medicine, New York, USA

Christopher I. Amos,
Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA

Margaret R. Spitz,
Dan L Cuncan Cancer Center, Baylor College of Medicine, Houston, USA

Chu Chen,
Fred Hutchinson Cancer Research Center, Seattle, USA

Gary Goodman,
Fred Hutchinson Cancer Research Center, Seattle, USA

Joachim Heinrich,
Institute of Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany

Heike Bickeböller,
Department of Genetic Epidemiology, University of Göttingen Medical School, Göttingen, Germany

Albert Rosenberger,
Department of Genetic Epidemiology, University of Göttingen Medical School, Göttingen, Germany

Angela Risch,
Division of Epigenomics and Cancer Risk Factors, Translational Lung Research Centre Heidelberg (TLRC-H), German Cancer Research Center, Heidelberg, Germany

Thomas Muley,
Division of Epigenomics and Cancer Risk Factors, Translational Lung Research Centre Heidelberg (TLRC-H), German Cancer Research Center, Heidelberg, Germany

© Springer-Verlag Berlin Heidelberg 2013

Correspondence to: Rayjean J. Hung, rayjean.hung@lunenfeld.ca.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-013-1270-y) contains supplementary material, which is available to authorized users.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

John R. McLaughlin,
Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON, Canada

Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Simone Benhamou,
INSERM U946, Fondation Jean Dausset-CEPH, Paris, France

Christine Bouchardy,
Geneva Cancer Registry, Geneva, Switzerland

Juan Pablo Lewinger,
University of Southern California, Los Angeles, USA

John S. Witte,
University of San Francisco, San Francisco, USA

Gary Chen,
University of Southern California, Los Angeles, USA

Shelley Bull, and
Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON, Canada

Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Rayjean J. Hung
Samuel Lunenfeld Research Institute of Mount Sinai Hospital, Toronto, ON, Canada

Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

Darren R. Brenner: brenner.darren@gmail.com; Rayjean J. Hung: rayjean.hung@lunenfeld.ca

Abstract

Recent evidence suggests that inflammation plays a pivotal role in the development of lung cancer. In this study, we used a two-stage approach to investigate associations between genetic variants in inflammation pathways and lung cancer risk based on genome-wide association study (GWAS) data. A total of 7,650 sequence variants from 720 genes relevant to inflammation pathways were identified using keyword and pathway searches from Gene Cards and Gene Ontology databases. In Stage 1, six GWAS datasets from the International Lung Cancer Consortium were pooled (4,441 cases and 5,094 controls of European ancestry), and a hierarchical modeling (HM) approach was used to incorporate prior information for each of the variants into the analysis. The prior matrix was constructed using (1) role of genes in the inflammation and immune pathways; (2) physical properties of the variants including the location of the variants, their conservation scores and amino acid coding; (3) LD with other functional variants and (4) measures of heterogeneity across the studies. HM affected the priority ranking of variants particularly among those having low prior weights, imprecise estimates and/or heterogeneity across studies. In Stage 2, we used an independent NCI lung cancer GWAS study (5,699 cases and 5,818 controls) for in silico replication. We identified one novel variant at the level corrected for multiple comparisons (rs2741354 in *EPHX2* at 8q21.1 with p value = 7.4×10^{-6}), and confirmed the associations between *TERT* (rs2736100) and the *HLA* region and lung cancer risk. HM allows for prior knowledge such as from bioinformatic sources to be incorporated into the analysis systematically, and it represents a complementary analytical approach to the conventional GWAS analysis.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Introduction

Epidemiologic evidence suggests that chronic severe inflammation may be related to carcinogenesis of the lung potentially through common exposures (infectious agents, particulate matter, smoke, fumes and exhausts) (Engels 2008), tumor initiation and promotion (Peek et al. 2005; Bernatsky et al. 2008; Parikh-Patel et al. 2008), as well as genetic determinants (Engels et al. 2007). Three recent investigations completed comprehensive analyses of genes involved in inflammation pathways and lung cancer risk based on GWAS data. Shi et al. (2012) investigated variants in inflammation pathways and computed gene-based association scores. Spitz et al. (2012a, b) examined variants from an inflammation panel of variants among never smokers using a two-stage approach and among current and former smokers using a three-stage approach, respectively. All three investigations identified novel variants (in *RAD52*, *NR4A1* and *IL2RB* and *BCL2L14* genes, respectively) using their approaches. Although these analyses chose variants based on their plausible biological function, neither systematically incorporated functional information into their analyses.

Given that highly significant variants for lung cancer risk have been identified through standard GWAS analysis using maximum likelihood (ML) approaches for single variants, the current challenge is how to identify the variants that might not reach GWAS level significance while still being biologically important. Hierarchical models/modeling (HM) presents an alternative for addressing some of the shortcomings of standard GWAS analysis by incorporating the wealth of readily available bioinformatic data characterizing the structural and functional roles of common variants (Cantor et al. 2010; Wang et al. 2010). The goal of HM in this application is to systematically incorporate available prior biological knowledge, improve effect and variance estimation in genomic investigations(Aragaki et al. 1997), and optimize variant prioritization for follow-up investigation (Witte and Greenland 1996; Witte 1997). A recent simulation study showed that an empirical-Bayes hierarchical framework outperforms traditional ML methods (increased power, reduced false-positive rate) and may suggest additional regions of interest beyond traditional ML methods (Heron et al. 2011).

We applied two HM methods developed for GWAS level data to optimize variant prioritization based on prior biological information. One model, developed by Chen and Witte (2007) estimates the effect of variants based on a single-distribution of variant effects. The other, developed by Lewinger et al. (2007) re-ranks variants assuming a two-distribution model, where the majority of the variant effects are centered at null and a small fraction of variant effects centered at a non-null value. We applied a two-stage design to investigate the genes in inflammation-related pathways: in Stage 1, we applied each of the two HM frameworks to pooled data from six lung cancer GWAS examining common variants from the International Lung Cancer Consortium (ILCCO); in Stage 2, we conducted in silico replication based on the DCEG lung cancer GWAS data.

Methods

Study populations

Within ILCCO (http://ilcco.iarc.fr) six case–control studies from Europe and North America participated in this investigation. All the studies were, at a minimum, frequency-matched based on age and sex. The subjects were all European descendants as described in the previous publications (Hung et al. 2008; Brenner et al. 2010; Brennan et al. 2006; Amos et al. 2008; Thornquist et al. 1993; Sauter et al. 2008). The combined population consisted of 4,441 cases and 5,194 controls. Additional study-specific details are summarized in Table 1. To further assess the performance of HM and the robustness of the findings, we used the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

DCEG lung cancer GWAS results available through the Database of Genotypes and Phenotypes (dbGAP) from the National Institutes of Health (NIH) to perform in silico replication of variants of interest identified by the two HM approaches. The DCEG lung cancer GWAS data (NCI-replication) consist of 506,062 variants from the 550 K Illumina chip on 5,699 cases and 5,818 controls recruited in four studies by National Cancer Institute (NCI) investigators (referred to as NCI-replication). The details of this study have been described previously (Landi et al. 2009).

Gene and variant selection

In order to identify relevant genes of interest, we used two electronic databases, Gene Cards (genecards.org) and Gene Ontology (geneontology.org) to search for inflammation-related genes. Specifically, keyword searches using “inflammation” or “lung specific inflammation” or “lung inflammation” in Gene Cards and the three Gene Ontology headings “inflammation”, “immune response” and “signaling” identified genes that are implicated in these particular pathways. A total of 720 genes were identified through the searches. This list was merged with annotation data from the Illumina Human-Hap 300, which includes 317 K SNP variants. Variants in the genes of interest that were present on the chip were mapped to the closest genes as allocated based on the data from Hapmap (The International HapMap Project 2003). This provided a list of 7,650 variants available for analysis.

Maximum likelihood estimation of variant effects

The input data for HM analyses were variant-specific pooled-effect estimates. All effect estimates from each study and pooled estimates were based on log-additive models and represent per allele ORs. In order to obtain these data across studies, study-specific log odds ratios were estimated for each variant from each of the six participating studies using logistic regression methods (PLINK, v1.07) (Purcell et al. 2007). To account for differences across study populations, we used random effects model to combine the effect estimates across studies, and p values for heterogeneity were estimated for each variant based on Q statistics (STATA v10, College Station, TX, USA).

Z-matrix formulation

One of the major strengths of the HM approach is that prior knowledge of biological function and genomic properties can be incorporated into effect estimation for the genetic variants of interest. The Z-matrix, a key component of the HM approach, was developed using information from several bioinformatic sources. In formulating the Z-matrix, we included both gene and variant-specific columns reflecting the levels and types of data included as described below. An example of the Z-matrix used with chosen column headings and the explanation of the scores is provided in Supplementary Table 1.

Gene-specific information

Gene-specific columns were created based on the function of the genes using the biological process sub-ontologies within the Gene Ontology (major headings listed above). Pathway/process scores were calculated based on the involvement of a gene in a pathway/ontology and the related sub-processes within that pathway to reflect how heavily involved a gene is in a particular function. For example, if a gene was involved in four sub-processes/ontologies within the inflammation response pathway (e.g. acute inflammation response, chronic inflammation, healing during inflammatory response, regulation of immune response), each of the variants in this gene was given a score of 4. Variables were created for each of the three ontologies of interest (inflammation, immune response, signaling).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Variant-specific information

Variant-specific columns were created based on PhastCons scores (http://www.genome.ucsc.edu), the SIFT score (http://sift.jcvi.org/), whether the variant is in an intergenic region, whether it is an exonic coding variant or a non-synonymous variant and its location relative to the start of the gene. The p value for heterogeneity across the studies from the pooling of the variant effect estimates was also included in the Z-matrix. This column was included to incorporate the aspect of pooled analysis into the HM framework as the input estimates for the variants are from a meta-analysis of studies. A variant-specific total (calculated as the sum of all Z-matrix columns) was calculated for each variant in order to determine the effects of the Z-matrix on the HM estimates. In principle, this approach suggests that variants with a higher a priori function in inflammatory responses and stronger genomic functionality will be given higher weight in analysis.

We also created a second Z-matrix to test the sensitivity of the models to variations in the Z-matrix and to include additional covariates to account for the available variants in linkage distribution (LD) with our tag variants. We incorporated information for those variants available from the 1000 Genomes Project Pilot I (Siva 2008) that were in LD (pairwise $R^2 > 0.8$ and within 500 kb) with the tag variants available from our 317 K data using the SNP Annotation and Proxy Search tool (SNAP) V2.2 from the Broad Institute (Johnson et al. 2008). We then obtained location data from GeneCruiser (Broad Institute) to include Z-matrix columns for the number of variants in LD with our genotyped variants that are in 3′ or 5′ untranslated regions (UTR), coding variants and whether the variants are non-synonymous. The number of directly genotyped variants and the number of variants in LD with the genotyped variants in each category are summarized in Supplementary Table 2. The same Z-matrices were used for both HM approaches.

Model specification for hierarchical modeling

We chose to apply two different HM models to the pooled ML data. The two methods (Lewinger et al. 2007; Chen and Witte 2007) have distinct differences although they are similar in goal and computation. The model of Chen and Witte provides for a single distribution of effects of the variants and uses a weighting variable in addition to the Z-matrix to further emphasize those believed more strongly a priori to be causal (henceforth referred to as the *one-distribution model*). In contrast, the model of Lewinger et al. accounts for two-prior distributions with GWAS level data suggesting a prior model for the true noncentrality parameters of variant associations composed of a large mass at zero and a continuous distribution of nonzero values (henceforth referred to as the *two-distribution model*). Complete statistical descriptions of the two models have been published previously (Lewinger et al. 2007; Chen and Witte 2007).We therefore provide only brief summaries of the modeling parameters used in the supplementary methods section.

Results comparison

We took two approaches to account for multiple testing. First, a Bonferroni correction was applied to our results for each variant whereby variants achieving significance at a level of $\alpha/7{,}650$ were considered noteworthy. This considered each variant comparison as independent. It is likely that this level of correction is too stringent as the test statistics are correlated due to nearby variants being in LD. Second, in order to reduce possible type II error, we used the methods of Gao et al. (2008) which account for the correlations between variants using composite LD across variant pairs to determine the effective number of comparisons for adjustment. This procedure suggested that the effective number of comparisons was $\alpha/4{,}603$ [multiple correction (MC) level significance]. To compare results from pooled ML to HM, we also examined those variants with test statistics exceeding critical values of $p < 0.05$ as well as <0.001.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

In order to determine the effects of the one-distribution model on effect estimates and priority, variants were ranked based on *p* values from pooled ML and HM. For the two-distribution model, the variants were ranked based on their posterior probability of association. This was suggested to be the most powerful ranking strategy of the three posterior statistics provided in the original methods paper (Lewinger et al. 2007) based on the model parameters used.

In silico replication

For variants with HM *p* value < 0.05 based on the one-distribution model or ranked in the top 100 based on the two-distribution model, we conducted in silico replication using the DCEG lung cancer GWAS. The analyses based on DCEG data were adjusted for age, sex, smoking groups and the top principal component of ethnicity because the population was the amalgamation of four independent studies. For the top variants of interest, we examined for the presence of differential effects by histology groups (adenocarcinoma, squamous cell carcinoma, large cell carcinoma), smoking groups (ever vs. never), gender and age at diagnosis (< 50 vs. ≥ 50 years of age) where possible in the original six studies and in the NCI-replication set.

Results

Pooled maximum-likelihood results from first stage model

There was no evidence of population structure within each of the populations [study-specific lambda inflation factor (λ): Toronto = 1.04, MDACC = 1.01, CARET = 1.06, CE = 1.06, Germany = 1.04, France = 1.03] and the overall λ for pooled *p* values was 1.012 when examining the bottom 90 % of the distribution. The absolute values of pooled ML βs ranged from 2.8×10^{-6} to 1.09 (corresponding to an OR range of 1.00–2.98). The proportion of variants that showed significant heterogeneity in effect estimates across studies at $p < 0.05$ was as expected (5.3 %). After pooling the results across the studies, 17 variants were significant at a level of $p < 0.001$, including two that were significant at MC level significance ($p < 6.54 \times 10^{-5}$) and one at Bonferroni corrected levels ($p < 1.09 \times 10^{-6}$) and standard GWAS level significance ($p < 5 \times 10^{-8}$) (Table 2).

Hierarchical modeling results

When comparing pooled ML to HM estimates, the amount of change can be described as a combination of the weight given in the Z-matrix, the standard error (SE) of the pooled estimate (Supplementary Figure 1) and the *p* value for heterogeneity (Supplementary Figure 2). The changes in ranks were altered by several of the columns of the Z-matrix with the largest effect being from the *p* value for heterogeneity column. Increased heterogeneity (small *p* value for heterogeneity) on average decreased the priority from ML to HM estimates. Therefore, variants with inconsistent effects across studies ranked on average at a decreased level of relative importance in HM estimates and those with consistent effects were modified heavily based on their Z-matrix and ML effects. As the first stage, SE and the *p* value for heterogeneity were correlated and similar effects were observed for changes in both measures.

One-distribution model

We observed 311 variants that were significant at a level of $p < 0.05$ in HM estimates. Among these, 15 variants were significant at $p < 0.05$ in HM but not in pooled ML. The 15 variants had an average higher Z-column score, lower first stage SE and tended to be very homogeneous across the studies compared to the distribution across all variants. On the other hand, 49 variants had ML *p* value < 0.05 but were no longer significant with HM.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

When examining those 49 variants, their first stage SE was higher compared to the total population of variants.

Two-distribution model

Result rankings changed more drastically between pooled ML and HM in the two-distribution model compared to the one-distribution model. The two-distribution model was sensitive to Z-matrix prior scores, particularly where variants showed small effect estimates. The two top-identified variants maintained their top rankings in both one-and two-distribution models. As the posterior probability from the two-distribution model does not follow the same distributional assumptions as the first stage p values, no direct comparison with regards to "significance" criteria of the p values from pooled ML estimates to HM estimates was possible. The two-distribution model, however, suggested only the top hit (rs2736100–*TERT*) to have a posterior probability of association to be <0.05 and that all others would not be suggested for further analysis. The top 100 variants from the two-distribution model were, however, included in the replication regardless of their posterior probability to assess whether the two-distribution identified any additional variants that were not captured in the standard GWAS analysis.

In silico replication

Of the 311 variants observed at $p < 0.05$ from HM from the one-distribution model, 32 were significantly replicated in the NCI-replication dataset at the nominal p value of < 0.05 including three replicated at the p value of < 0.001. For the top 100 variants identified from the two-distribution HM, five were replicated and these were already included in the 32 replicated from the one-distribution model.

Of the 32 replicated variants, 16 were in the 5p15 and 6p21–22 regions which have been previously highlighted in lung cancer GWAS (Hung et al. 2008; Landi et al. 2009), including the variant significant at GWAS levels (rs2736100) in both HM and pooled ML. The remaining 16 variants represented 10 independent regions of interest, which includes one novel variant, rs2741354 located on chromosome 8q21 near the epoxide hydrolase-2 (EPHX2) gene. This variant did not reach significance at MC level in the Discovery set based on conventional ML estimate; however, it did convey significance based on HM estimate. When pursuing in silico replication, the variant became significant at MC level ($p = 7.38 \times 10^{-6}$) when combining the data of six ILCCO studies and DCEG lung cancer study. The other variant that was significant at MC levels in pooled ML (rs1023253) was not found in the NCI-replication data, however, a proxy variant ($r^2 = 0.972$ from HapMap CEU) rs6424779 was examined and effects were not replicated ($p = 0.3214$). Table 3 contains the 32 variants that were replicated in the NCI-replication dataset. A forest plot for the top novel signal at ch8p21.1 (rs2741354) including subgroup analyses by histology, smoking, sex and age at onset is shown in Fig. 1. No specific effect modification by age, smoking status, gender or histology was observed.

Discussion

In this investigation into the role of variants in inflammation-related pathways using HM, we confirmed the associations with TERT and HLA regions. We also identified a novel locus in EPHX2 at 8p21.1, which would not have been identified by the conventional ML approach with multiple comparison adjustment. Our results in concert with others (Heron et al. 2011) suggest that strong true-positive variants would not be missed from the use of HM in a GWAS scale analysis, whereas the incorporation of the prior knowledge in a systematic manner can help to identify novel variants that do not reach GWAS significance level.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We observed an association between lung cancer risk and variant rs2741354 on chromosome 8p21.1 situated downstream from the epoxide hydrolase-2 (*EPHX2*) gene which is expressed in lung tissue samples (Yanai et al. 2005). This variant, which is not in a coding region was included based on its role in the inflammation pathway from Gene Ontology and had consistent effects across studies (p value for heterogeneity = 0.85). To our knowledge, this variant has not been implicated in lung cancer risk to date; however, other epoxide hydrolase genes have been associated with lung cancer risk in candidate studies (Lee et al. 2002). Epoxide hydrolases play an important role in the lung by metabolizing inhaled irritants and carcinogens (Petruzzelli et al. 1992), and variants in the genes have been shown to be related to other inflammatory disease risk (Korotkova et al. 2011). We examined the effects of the variant across smoking, gender, age at onset and histology groups to determine whether the variant was acting as a marker of exposure or subgroup risk. This did not seem to be the case as the combined effects across groups were consistent although we were relatively limited in our analyses of never smokers as only two studies possessed never smoking cases (Fig. 1).

Our analyses further validated the finding of variants in the *TERT* gene (rs2736100) and its relationship with lung cancer risk. The variant under investigation has been previously implicated in lung cancer risk in GWAS analysis (Hung et al. 2008) and targeted replication (McKay et al. 2008). The *TERT* gene was included into the analysis based on the results from a Gene Cards search for inflammation-related genes. *TERT* is thought to be active in some rapidly dividing cells of the immune system. It is also believed to be related to endothelial nitric oxide synthase control (Matsushita et al. 2001). Both exogenously and endogenously-induced oxidative stress leads to translocation of human *TERT* from the nucleus into the cytosol (Santos et al. 2004). Mutations in *TERT* have been related to idiopathic pulmonary fibrosis (Armanios et al. 2007). Fibrotic lung scaring has also been associated with lung cancer in prospective observational studies (Yu et al. 2008), and other inflammatory lung diseases have been consistently associated with lung cancer risk (Brenner et al. 2011).

Through the use of HM, we identified 15 variants at significance levels of $p < 0.05$ not observed using pooled ML estimates, two of which were replicated in the NCI-replication data. This proportion is greater than expected by chance (0.75/15 significant tests at $\alpha = 0.05$) and these variants would have been missed without the use of HM. One of the variants is of particular interest (rs3129871) as it is found in the Human Leukocyte Antigen (*HLA-DRA*) gene on chromosome 6p21. This area has been previously associated with lung cancer in a previous linkage study (Bailey-Wilson et al. 2004) and in previous GWAS analyses (Landi et al. 2009). Variants in several *HLA* genes (*HCG9*, *HLA-B*, *HLA-A*) also ranked 4, 8, and 9th in the overall pooled ML rankings and were robust in HM analysis. Supplementary Figure 3 displays the significance of variants from our initial discovery set across the *HLA* region. Our results corroborate a previous GWAS (Wang et al. 2008) that suggests the possibility of multiple independent variants in the region. It remains an item of debate as to whether this area represents a marker of risk or regional population substructure (Landi et al. 2009). Nevertheless, the crucial role of *HLA* in the inflammation and immune responses is well documented (McDevitt 1980) and the association with lung cancer provides additional weight to the inflammatory origin of lung cancer.

Concerning methodology, we found the HM approach to be complementary to the standard GWAS analysis. With informative prior data, this type of analytical approach can account for functional information in a systematic manner, instead of simply as ad hoc interpretation. By systematically including the data into the analysis, the method aids in the identification of true associations. Thus, the biological knowledge that is traditionally consulted after the discovery phase of top hits is systematically incorporated in the analysis. Moreover,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

previous HM analyses suggest that even completely uninformative information will not reduce the power of the analysis beyond traditional methods (Heron et al. 2011). In this application, we consider HM not as a simple ranking tool, but instead as a valuable alternative analytical approach to help identify variants with potential biological function that could have been missed in the standard GWAS analysis.

We propose the inclusion of *p* values for heterogeneity as a weight function column when applying the method to pooled multi-center data as an extension of the methods proposed in the paper by Chen and Witte (2007). Even minimal heterogeneity across study populations as observed in this case is a determinant of the distribution of pooled ML effects as input data. Failure to account for this design feature may lead to biased results as the distribution of pooled effects may not reflect the underlying distributions of variant effects.

There are several limitations in this analysis. We employed random effects model to pool effect estimates across study sites. This method is conservative (Thompson et al. 2011) and despite the populations being of European decent, we felt it could be inappropriate to utilize fixed effect models to pool study-specific estimates because of differences across populations in exposures as well as different covariate assessment. Consequently, standard errors in pooled ML estimates may have been overestimated for some variants and altered the relative change in rank from first to second stage estimates. We used those genes with direct coverage on the Illumina 317 K chip. It is indeed probable that there are other known and uncharacterized genes and variants with effects on lung cancer through their role in the inflammatory response that were omitted based on the selection process. In addition, the variants chosen for inclusion on the platforms were not chosen based on function, but rather on tagging ability from HapMap data (Illumina 2006, 2010) and therefore may not provide a representative sample of functional variants within the genes. In order to address this shortcoming, one of the formulations of the Z-matrix as utilized included additional information whereby variants in linkage disequilibrium with other functional variants in the 1000 Genomes Project data were given additional weight as proposed in the methods of Heron (Heron et al. 2011). The evaluation by Witte and Chen, however, showed that the model and the subsequent rankings provided by HM were relatively robust to large variations in the Z-matrix formulation provided that the reduced Z-matrix remains informative despite alterations (Chen and Witte 2007). We observed a similar effect whereby the relative rankings of variants were quite robust to additions and variations to the Z matrix.

Our HM analysis was conducted across all cases and subgroups combined. It may be useful to conduct additional HM analyses by smoking and other subgroups as the differential effects of the inflammatory pathways on lung carcinogenesis remain unclear at this point. In this case as we did not have individual-level data for all studies to be able to conduct full subgroup analyses and were greatly limited in power for certain subgroups, in particular never smokers, we only conducted HM among the overall results and examined for differential effects among top hits where possible across subgroups.

Conclusion

In conclusion, we extended previously developed methods using HM in a targeted pathway-specific approach to pooled GWAS data for inflammation-related genes. The added values of hierarchical modeling are (1) to identify potential new loci that would have been missed in conventional analysis, and (2) reduce false positives by incorporating priors in the analysis stage instead of use it as ad hoc interpretation. In addition to confirming the previous known loci, we identified a novel locus at 8p21 in the EPHX2 gene not previously identified and worthy of follow-up investigation. Given that the highly significant variants

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

can be and have been detected by the standard GWAS analysis, use of analytical methods that incorporate prior information would be one of the most cost-effective approaches to uncover the susceptibility loci that do not reach GWAS level significance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

RJH holds a Cancer Care Ontario Chair in Population Studies. DRB holds a Canadian Institutes of Health Research Canada Graduate Scholarship. This research was supported by funding from Training grant GET-101831. DRB is a fellow of CIHR STAGE (Strategic Training for Advanced Genetic Epidemiology) – CIHR Training Grant in Genetic Epidemiology and Statistical Genetics. The work is supported by a Canadian Cancer Society Research Institute grant (no. 020214) and a U19 grant from the National Institutes of Health (U19 CA148127). The MD Anderson study was supported by a grant from the National Institutes of Health CA127219. The central Europe study was a multi-center study conducted in seven central European countries. The following investigators are responsible for the collection of data at each of the sites: Neonila Szeszenia-Dabrowska, Jolanta Lissowska, David Zaridze, Peter Rudnai, Eleonora Fabianova, Lenka Foretova, Vladimir Janout, Vladimir Bencko, Miriam Schejbalova.

References

Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008; 40(5):616–622. [PubMed: 18385676]

Aragaki CC, Greenland S, Probst-Hensch N, Haile RW. Hierarchical modeling of gene-environment interactions: estimating NAT2 genotype-specific dietary effects on adenomatous polyps. Cancer Epidemiol Biomarkers Prev. 1997; 6(5):307–314. [PubMed: 9149889]

Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA 3rd, Lansdorp PM, Greider CW, Loyd JE. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med. 2007; 356(13):1317–1326. [PubMed: 17392301]

Bailey-Wilson JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, Wiest JS, Fain P, Schwartz AG, You M, Franklin W, Klein C, Gazdar A, Rothschild H, Mandal D, Coons T, Slusser J, Lee J, Gaba C, Kupert E, Perez A, Zhou X, Zeng D, Liu Q, Zhang Q, Seminara D, Minna J, Anderson MW. A major lung cancer susceptibility locus maps to chromosome 6q23-25. Am J Hum Genet. 2004; 75(3):460–474. [PubMed: 15272417]

Bernatsky S, Clarke AE, Ramsey-Goldman R. Cancer in systemic lupus: what drives the risk? Cancer Causes Control CCC. 2008; 19(10):1413–1414. [PubMed: 18575952]

Brennan P, Crispo A, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Fletcher T, Boffetta P. High cumulative risk of lung cancer death among smokers and nonsmokers in Central and Eastern Europe. Am J Epidemiol. 2006; 164(12):1233–1241. [PubMed: 17032696]

Brenner DR, Hung RJ, Tsao MS, Shepherd FA, Johnston MR, Narod S, Rubenstein W, McLaughlin JR. Lung cancer risk in never-smokers: a population-based case–control study of epidemiologic risk factors. BMC Cancer. 2010; 10:285. [PubMed: 20546590]

Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS One. 2011 (accepted).

Cantor RM, Lange K, Sinsheimer JS. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. Am J Hum Genet. 2010; 86(1):6–22. [PubMed: 20074509]

Chen GK, Witte JS. Enriching the analysis of genome wide association studies with hierarchical modeling. Am J Hum Genet. 2007; 81(2):397–404. [PubMed: 17668389]

Engels EA. Inflammation in the development of lung cancer: epidemiological evidence. Expert Rev Anticancer Ther. 2008; 8(4):605–615. [PubMed: 18402527]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Engels EA, Wu X, Gu J, Dong Q, Liu J, Spitz MR. Systematic evaluation of genetic variants in the inflammation pathway and risk of lung cancer. Cancer Res. 2007; 67(13):6520–6527. [PubMed: 17596594]

Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008; 32(4):361–369. [PubMed: 18271029]

Heron EA, O'Dushlaine C, Segurado R, Gallagher L, Gill M. Exploration of empirical Bayes hierarchical modeling for the analysis of genome-wide association study data. Biostatistics. 2011; 12(3):445–461. [PubMed: 21252078]

Hung RJ, Baragatti M, Thomas D, McKay J, Szeszenia-Dabrowska N, Zaridze D, Lissowska J, Rudnai P, Fabianova E, Mates D, Foretova L, Janout V, Bencko V, Chabrier A, Moullan N, Canzian F, Hall J, Boffetta P, Brennan P. Inherited predisposition of lung cancer: a hierarchical modeling approach to DNA repair and cell cycle control pathways. Cancer Epidemiol Biomarkers Prev. 2007; 16(12):2736–2744. [PubMed: 18086781]

Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcatova I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature. 2008; 452(7187):633–637. [PubMed: 18385738]

Illumina. Tecnhical Bulletin: whole-genome genotyping with the Sentrix HumanHap300 Genotyping BeadCHip and the Infinium II Assay. San Diego: Illumina; 2006.

Illumina. The power of intelligent SNP selection: a technical note. San Diego: Illumina; 2010.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008; 24(24): 2938–2939. [PubMed: 18974171]

Korotkova M, Daha NA, Seddighzadeh M, Ding B, Catrina AI, Lindblad S, Huizinga TW, Toes RE, Alfredsson L, Klareskog L, Jakobsson PJ, Padyukov L. Variants of gene for microsomal prostaglandin E2 synthase show association with disease and severe inflammation in rheumatoid arthritis. Eur J Hum Genet. 2011; 19(8):908–914. [PubMed: 21448233]

Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, Vooder T, Nelis M, Valk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeboller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S, Caporaso NE. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009; 85(5):679–691. [PubMed: 19836008]

Lee WJ, Brennan P, Boffetta P, London SJ, Benhamou S, Rannug A, To-Figueras J, Ingelman-Sundberg M, Shields P, Gaspari L, Taioli E. Microsomal epoxide hydrolase polymorphisms and lung cancer risk: a quantitative review. Biomarkers. 2002; 7(3):230–241. [PubMed: 12141066]

Lewinger JP, Conti DV, Baurley JW, Triche TJ, Thomas DC. Hierarchical Bayes prioritization of marker associations from a genome-wide association scan for further investigation. Genet Epidemiol. 2007; 31(8):871–882. [PubMed: 17654612]

Matsushita H, Chang E, Glassford AJ, Cooke JP, Chiu CP, Tsao PS. eNOS activity is reduced in senescent human endothelial cells: preservation by hTERT immortalization. Circ Res. 2001; 89(9): 793–798. [PubMed: 11679409]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

McDevitt HO. Regulation of the immune response by the major histocompatibility system. New Eng J Med. 1980; 303(26):1514–1517. [PubMed: 6776404]

McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njolstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubinski J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, Gonzalez CA, Ramon Quiros J, Martinez C, Navarro C, Ardanaz E, Larranaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjonneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008; 40(12):1404–1406. [PubMed: 18978790]

Parikh-Patel A, White RH, Allen M, Cress R. Cancer risk in a cohort of patients with systemic lupus erythematosus (SLE) in California. Cancer Causes Control CCC. 2008; 19(8):887–894. [PubMed: 18386139]

Peek RM Jr, Mohla S, DuBois RN. Inflammation in the genesis and perpetuation of cancer: summary and recommendations from a national cancer institute-sponsored meeting. Cancer Res. 2005; 65(19):8583–8586. [PubMed: 16204020]

Petruzzelli S, Franchi M, Gronchi L, Janni A, Oesch F, Pacifici GM, Giuntini C. Cigarette smoke inhibits cytosolic but not microsomal epoxide hydrolase of human lung. Hum Exp Toxicol. 1992; 11(2):99–103. [PubMed: 1349227]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]

Santos JH, Meyer JN, Skorvaga M, Annab LA, Van Houten B. Mitochondrial hTERT exacerbates free-radical-mediated mtDNA damage. Aging Cell. 2004; 3(6):399–411. [PubMed: 15569357]

Sauter W, Rosenberger A, Beckmann L, Kropp S, Mittelstrass K, Timofeeva M, Wolke G, Steinwachs A, Scheiner D, Meese E, Sybrecht G, Kronenberg F, Dienemann H, Chang-Claude J, Illig T, Wichmann HE, Bickeboller H, Risch A. Matrix metalloproteinase 1 (MMP1) is associated with early-onset lung cancer. Cancer Epidemiol Biomarkers Prev. 2008; 17(5):1127–1135. [PubMed: 18483334]

Shi J, Chatterjee N, Rotunno M, Wang Y, Pesatori AC, Consonni D, Peng L, Wheeler W, Broderick P, Henrion M, Eisen T, Wang Z, Chen W, Dong Q, Albanes D, Thun M, Spitz MR, Bertazzi PA, Caporaso NE, Chanock SJ, Amos CI, Houlston RS, Landi MT. Inherited variation at chromosome 12p13.33, including RAD52, influences the risk of squamous cell lung carcinoma. Cancer Discovery. 2012; 2(2):131–139. [PubMed: 22585858]

Siva N. 1000 Genomes project. Nat Biotechnol. 2008; 26(3):256. [PubMed: 18327223]

Spitz M, Gorlov I, Dong Q, Wu X, Chen W, Chang D, Etzel C, Caporaso NE, Zhao Y, Christiani DC, Brennan P, Albanes D, Shi J, Thun MJ, Landi MT, Amos CI. Multistage analysis of variants in the inflammation pathway and lung cancer risk in smokers. Cancer Discovery. 2012; 3(1)

Spitz MR, Gorlov IP, Dong Q, Wu X, Chen W, Chang DW, Etzel CJ, Caporaso NE, Zhao Y, Christiani DC, Brennan P, Albanes D, Shi J, Thun M, Landi MT, Amos CI. Multistage analysis of variants in the inflammation pathway and lung cancer risk in smokers. Cancer Epidemiol Biomarkers Prev. 2012b; 21(7):1213–1221. [PubMed: 22573796]

The International HapMap Project. Nature. 2003; 426(6968):789–796. [PubMed: 14685227]

Thompson JR, Attia J, Minelli C. The meta-analysis of genome-wide association studies. Brief Bioinform. 2011; 12(3):259–269. [PubMed: 21546449]

Thornquist MD, Omenn GS, Goodman GE, Grizzle JE, Rosenstock L, Barnhart S, Anderson GL, Hammar S, Balmes J, Cherniack M, et al. Statistical design and monitoring of the Carotene and Retinol Efficacy Trial (CARET). Control Clin Trials. 1993; 14(4):308–324. [PubMed: 8365195]

Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008; 40(12):1407–1409. [PubMed: 18978787]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet. 2010; 11(12):843–854. [PubMed: 21085203]

Witte JS. Genetic analysis with hierarchical models. Genet Epidemiol. 1997; 14(6):1137–1142. [PubMed: 9433637]

Witte JS, Greenland S. Simulation study of hierarchical regression. Stat Med. 1996; 15(11):1161–1170. [PubMed: 8804145]

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, Lancet D, Shmueli O. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005; 21(5):650–659. [PubMed: 15388519]

Yu YY, Pinsky PF, Caporaso NE, Chatterjee N, Baumgarten M, Langenberg P, Furuno JP, Lan Q, Engels EA. Lung cancer risk following detection of pulmonary scarring by chest radiography in the prostate, lung, colorectal, and ovarian cancer screening trial. Arch Intern Med. 2008; 168(21): 2326–2332. (discussion 2332). [PubMed: 19029496]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

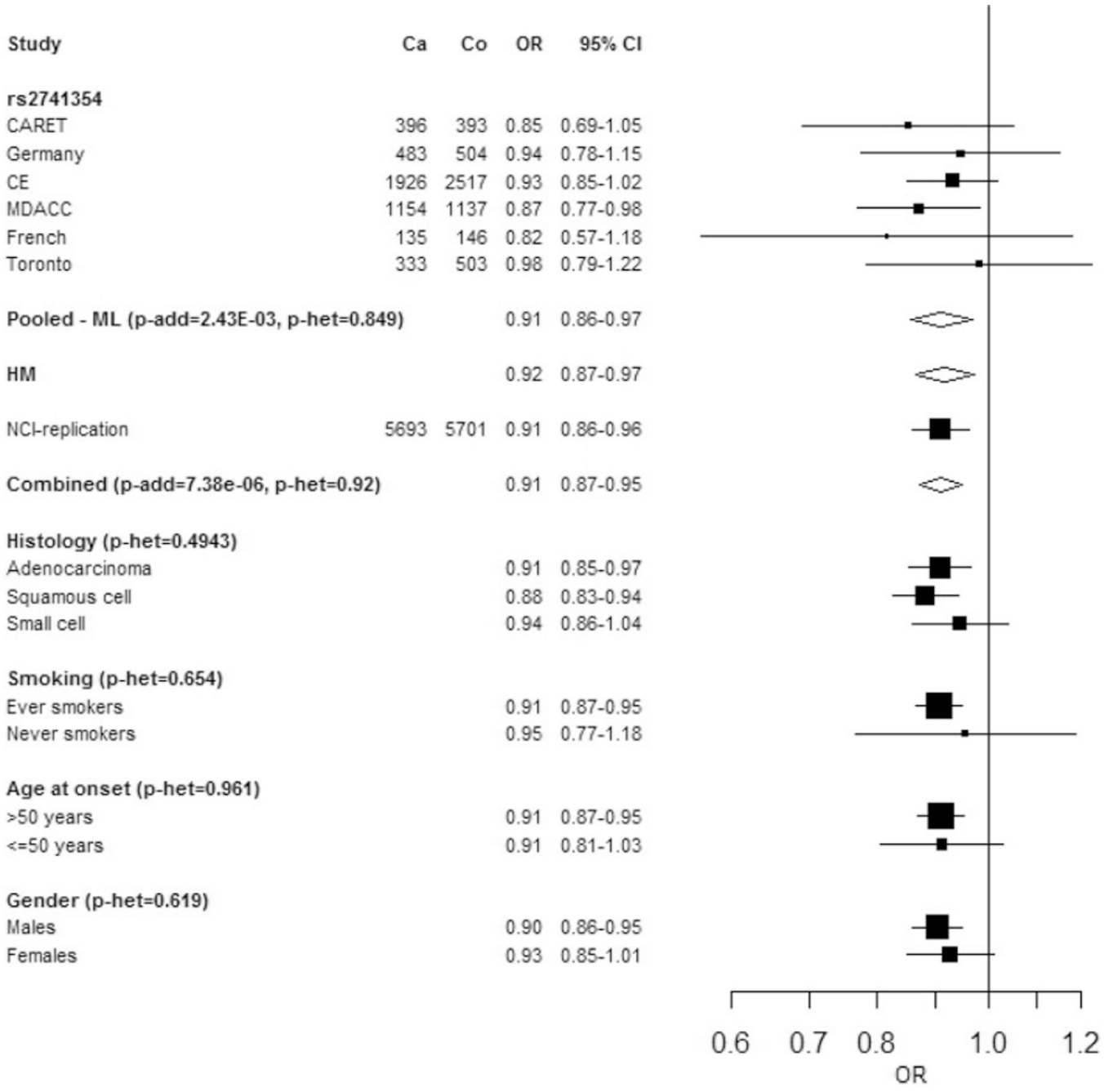

Fig. 1.

Forest plots of the top novel variant rs2741354 with subgroup analyses by histology, smoking, age at onset and gender. Pooled-ML refers to combined effects across all six studies. HM refers to estimates from the one-distribution model. NCI-replication refers to the in silico replication set. Effect estimates based on log-additive models and represent per allele ORs

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Descriptions of the studies and populations used in the analysis for pooled maximum likelihood estimates

Study	Location	Study design	Study feature	Study period	Platform	Case no.	Control no.
Central Europe	Europe	Hospital-based	None	1998–2002	HumanHap 317 K	1926	2,522
Toronto/SLRI	Canada	Hospital and Population-based	None	1997–2002	HumanHap 317 K	332	505
MDACC	USA	Hospital-based	Smokers	2001–2008	HumanHap 317 K	1154	1137
Germany	Europe	Population-based	Less than 55 years of age	1997–2004	HumanHap 550 K	497	491
CARET	USA	Nested case–control	Smokers	1983–1994	HumanHap 370 K Duo	397	393
France	Europe	Hospital-based	Smokers	1988–1992	HumanHap 370 K Duo	135	146

None denotes a sample with no defining characteristics reflecting the exposure distribution of the population sampled

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Numbers of variants meeting significance criteria from one-distribution HM and ML models

	ML	Uncorrected		MC[a]	Bonferroni	GWAS
HM	$p > 0.05$	$p < 0.05$	$p < 0.001$	$p < 6.54 \times 10^{-5}$	$p < 1.09 \times 10^{-6}$	$p < 5 \times 10^{-8}$
$p > 0.05$	7,290	49	0	0	0	0
$p < 0.05$	15	279	5	0	0	0
$p < 0.001$	0	0	9	1	0	0
$p < 6.54 \times 10^{-5}$	0	0	0	1	0	0
$p < 1.09 \times 10^{-6}$	0	0	0	0	0	0
$p < 5 \times 10^{-8}$	0	0	0	0	0	1

Each SNP contributes to only one cell

ML maximum likelihood, *HM* hierarchical modeling

[a]Multiple comparisons correction from Gao et al. (2008)

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Nominally significant ($p < 0.05$) variants from hierarchical modeling that replicated in NCI data

Variant	Chr	Pos	Gene	Risk allele	First stage					Replication				
					OR	L95	U95	*p* value (ML)	*p* value (HM)	OR	95 % CI		*p* value	Combined *p* value
rs10864368	1	8852579	*ENO1*	T	0.93	0.99	0.88	1.50×10^{-2}	1.58×10^{-2}	0.94	0.89	1.00	3.33×10^{-2}	1.14×10^{-3}
rs4674301	2	219068367	*SLC11A1*	C	0.92	0.98	0.86	9.38×10^{-3}	9.76×10^{-3}	0.92	0.87	0.98	8.33×10^{-3}	2.14×10^{-4}
rs3816560	2	219080347	*SLC11A1*	C	0.92	0.99	0.85	3.40×10^{-2}	3.22×10^{-2}	0.93	0.88	0.99	3.23×10^{-2}	5.18×10^{-3}
rs1027241	3	46287543	*CCR3*	A	1.08	1.16	1.00	4.00×10^{-2}	4.08×10^{-2}	1.06	1.01	1.13	3.07×10^{-2}	1.97×10^{-3}
rs222020	4	73001307	*GC*	C	1.09	1.19	1.00	3.97×10^{-2}	4.63×10^{-2}	1.08	1.01	1.17	3.55×10^{-2}	3.30×10^{-3}
rs4647992	4	103812532	*NFKB1*	T	1.22	1.40	1.05	7.33×10^{-3}	2.30×10^{-2}	1.15	1.01	1.32	3.81×10^{-2}	8.21×10^{-4}
rs2082317	4	172199279	AADAT	C	0.94	0.99	0.89	2.89×10^{-2}	2.91×10^{-2}	0.94	0.89	1.00	4.21×10^{-2}	2.88×10^{-3}
rs3775291	4	187379223	*TLR3*	T	1.08	1.15	1.01	2.28×10^{-2}	2.20×10^{-2}	1.08	1.02	1.15	9.35×10^{-3}	5.56×10^{-4}
rs2736100	5	1339516	*TERT*	C	1.22	1.29	1.15	5.56×10^{-11}	1.51×10^{-9}	1.10	1.03	1.16	1.30×10^{-3}	8.00×10^{-12}
rs2853676	5	1341547	*TERT*	T	1.08	1.16	1.01	2.82×10^{-2}	2.99×10^{-2}	1.08	1.02	1.15	1.11×10^{-2}	7.40×10^{-4}
rs2734986	6	29926547	*HLA-G*	C	1.17	1.30	1.06	2.92×10^{-3}	9.00×10^{-3}	1.11	1.03	1.21	8.62×10^{-3}	4.10×10^{-6}
rs2734985	6	29926641	*HLA-G*	C	1.11	1.19	1.03	5.43×10^{-3}	7.11×10^{-3}	1.11	1.03	1.18	4.80×10^{-3}	7.48×10^{-5}
rs2517861	6	29929961	*HLA-G*	T	1.10	1.18	1.02	9.99×10^{-3}	1.19×10^{-2}	1.11	1.03	1.19	3.81×10^{-3}	1.09×10^{-4}
rs2523946	6	30049922	*HCG9*	T	0.91	0.99	0.83	3.21×10^{-2}	3.84×10^{-2}	0.92	0.87	0.97	4.04×10^{-3}	1.57×10^{-5}
rs3132685	6	30053928	*HCG9*	A	1.22	1.35	1.11	4.60×10^{-5}	3.31×10^{-4}	1.14	1.03	1.26	8.93×10^{-3}	2.12×10^{-6}
rs3094694	6	30559883	*HLA-X10*	C	1.10	1.19	1.02	1.39×10^{-2}	1.74×10^{-2}	1.16	1.08	1.25	6.00×10^{-5}	3.91×10^{-6}
rs6457374	6	31380240	*HLA-C*	C	1.16	1.25	1.08	8.20×10^{-5}	2.40×10^{-4}	1.09	1.02	1.17	1.14×10^{-2}	2.93×10^{-6}
rs3873386	6	31381724	*HLA-C*	C	0.94	1.00	0.89	4.84×10^{-2}	4.10×10^{-2}	0.93	0.88	0.99	1.80×10^{-2}	2.09×10^{-3}
rs2523554	6	31439808	*HLA-B*	C	1.18	1.28	1.08	3.33×10^{-4}	1.20×10^{-3}	1.06	1.00	1.13	4.30×10^{-2}	8.24×10^{-7}
rs2428486	6	31462083	*MICA*	C	0.93	1.00	0.87	4.85×10^{-2}	4.04×10^{-2}	0.94	0.89	1.00	4.56×10^{-2}	2.44×10^{-3}
rs2523467	6	31470909	*MICB*	T	0.93	1.00	0.87	4.31×10^{-2}	3.58×10^{-2}	0.94	0.89	1.00	4.48×10^{-2}	3.07×10^{-3}
rs1131896	6	31487094	*MICA*	A	0.90	0.99	0.82	3.42×10^{-2}	4.18×10^{-2}	0.94	0.88	1.00	4.15×10^{-2}	1.18×10^{-3}
rs2844511	6	31497763	*MICA*	A	1.07	1.13	1.00	3.57×10^{-2}	3.53×10^{-2}	1.07	1.01	1.14	1.52×10^{-2}	5.73×10^{-3}
rs1052486	6	31718665	*BAT3*	A	1.07	1.14	1.00	4.70×10^{-2}	3.39×10^{-2}	1.06	1.00	1.12	4.70×10^{-2}	3.22×10^{-3}
rs3117582	6	31728499	*BAT3*	G	1.24	1.46	1.06	7.89×10^{-3}	4.25×10^{-2}	1.16	1.05	1.27	2.72×10^{-3}	3.87×10^{-7}
rs3129871	6	32514320	*HLA-DRA*	A	0.94	0.88	1.00	5.40×10^{-2}	3.92×10^{-2}	0.94	0.89	1.00	4.41×10^{-2}	5.32×10^{-3}

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Variant	Chr	Pos	Gene	Risk allele	First stage					Replication				
					OR	L95	U95	*p* value (ML)	*p* value (HM)	OR	95 % CI		*p* value	Combined *p* value
rs2187668	6	32713862	*HLA-DQA1*	T	1.21	1.35	1.09	5.61×10^{-4}	3.01×10^{-3}	1.12	1.03	1.22	1.10×10^{-2}	4.16×10^{-6}
rs2741354	8	27471393	*EPHX2*	C	0.91	0.97	0.86	2.43×10^{-3}	2.96×10^{-3}	0.91	0.86	0.96	9.60×10^{-4}	7.38×10^{-6}
rs11780592	8	27474664	*EPHX2*	G	0.92	0.99	0.84	3.15×10^{-2}	3.65×10^{-2}	0.90	0.83	0.97	4.81×10^{-3}	4.19×10^{-4}
rs10988496	9	129597055	*PTGES*	G	1.12	1.22	1.02	1.38×10^{-2}	2.04×10^{-2}	1.10	1.01	1.20	2.57×10^{-2}	9.11×10^{-4}
rs371352	15	67103257	*NOX5*	A	0.94	0.89	1.00	5.09×10^{-2}	4.19×10^{-2}	0.94	0.89	1.00	3.78×10^{-2}	4.37×10^{-3}
rs4243233	16	30439361	*ITGAL*	G	0.93	0.98	0.88	1.06×10^{-2}	9.26×10^{-3}	0.94	0.89	0.99	2.84×10^{-2}	7.96×10^{-4}

Estimates from original six studies were estimated using maximum likelihood and random effects model. Effect estimates based on log-additive models and represent per allele ORs. Variants on chromosome 5 and 6 have been previously reported (McKay et al. 2008; Landi et al. 2009; Wang et al. 2008)

Chr chromosome, *ML* maximum likelihood, *HM* hierarchical model