BRIEF COMMUNICATION

Pleiotropic Analysis of Lung Cancer and Blood Triglycerides

Verena Zuber, Crystal N. Marconett, Jianxin Shi, Xing Hua, William Wheeler, Chenchen Yang, Lei Song, Anders M. Dale, Marina Laplana, Angela Risch, Aree Witoelar, Wesley K. Thompson, Andrew J. Schork, Francesco Bettella, Yunpeng Wang, Srdjan Djurovic, Beiyun Zhou, Zea Borok, Henricus F. M. van der Heijden, Jacqueline de Graaf, Dorine Swinkels, Katja K. Aben, James McKay, Rayjean J. Hung, Heike Bikeböller, Victoria L. Stevens, Demetrius Albanes, Neil E. Caporaso, Younghun Han, Yongyue Wei, Maria Angeles Panadero, Jose I. Mayordomo, David C. Christiani, Lambertus Kiemeney, Ole A. Andreassen, Richard Houlston, Christopher I. Amos, Nilanjan Chatterjee, Ite A. Laird-Offringa, Ian G. Mills*, Maria Teresa Landi*

Affiliations of authors: NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine (VZ, AW, FB, YW, OAA), and Prostate Cancer Research Group, Centre for Molecular Medicine Norway, Nordic EMBL Partnership (VZ, IGM), University of Oslo, Oslo, Norway; Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway (VZ, AW, FB, YW, OAA); European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK (VZ); Departments of Surgery and of Biochemistry and Molecular Biology (CNM, CY, IALO) and Will Rogers Institute Pulmonary Research Center and Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine (BZ, ZB), Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA (CNM, CY, IALO); Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, US Public Health Service, Bethesda, MD (JS, XH, LS, DA, NEC, NC, MTL); Information Management Services, Inc.; Rockville, MD (WW); Multimodal Imaging Laboratory (AMD, AJS), Center for Human Development (AMD, AJS), Department of Radiology (AMD), Department of Neurosciences (AMD), Department of Psychiatry (WKT), and Cognitive Sciences Graduate Program (AJS), University of California, San Diego, La Jolla, CA; Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany (ML, AR); Department of Molecular Biology, University of Salzburg, Salzburg, Austria (AR); Translational Lung Research Center Heidelberg TLRC-H, Member of the German Center for Lung Research DZL, Heidelberg, Germany (AR); Department of Medical Genetics (SD) and Department of Molecular Oncology, Institute of Cancer Research and Department of Urology (IGM), Oslo University Hospital, Oslo, Norway (SD); NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway (SD); Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, the Netherlands (HFMvdH, JdG); Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands (DS); Netherlands Comprehensive Cancer Organization, Utrecht, the Netherlands (KKA); International Agency for Research on Cancer (IARC, World Health Organization [WHO]), Lyon, France (JM); Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada (RJH); Zentrum Informatik, Statistik und Epidemiologie, Universitätsmedizin Göttingen (HB); Epidemiology Research Program, American Cancer Society, Atlanta, GA (VLS); Center for Genomic Medicine, Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH (YH, CIA); Department of Epidemiology and Environmental Health, Harvard School of Public Health, Boston, MA (YW, MAP, DCC); Division of Medical Oncology, Ciudad de Coria Hospital, Coria, Spain (JIM); Division of Medical Oncology, University Hospital, Zaragoza, Spain (LK); Division of Pulmonary/Critical Care, Department of Medicine Massachusetts General Hospital/Harvard Medical School, Boston, MA (DCC); Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK (RH); Prostate Cancer UK/Movember Centre of Excellence for Prostate Cancer Research, Centre for Cancer Research and Cell Biology, Queen's University, Belfast, UK (IGM).

Correspondence to: Maria Teresa Landi, M.D., Ph.D., Senior Investigator, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, 9609 Medical Center Drive, Room 6E446, MSC 9769, Bethesda, MD 20892-9769 (e-mail: landim@mail.nih.gov) and Ian Geoffrey Mills, M.Biochem. , Ph.D., Centre for Cancer Research and Cell Biology (CCRCB), Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, United Kingdom (e-mail: I.Mills@qub.ac.uk), Prostate Cancer UK/Movember Centre of Excellence for Prostate Cancer Research, Centre for Cancer Research and Cell Biology, Queen's University, Belfast, UK.

*Authors contributed equally to this work.

Abstract

Epidemiologically related traits may share genetic risk factors, and pleiotropic analysis could identify individual loci associated with these traits. Because of their shared epidemiological associations, we conducted pleiotropic analysis of genome-wide association studies of lung cancer (12 160 lung cancer case patients and 16 838 control subjects) and cardiovascular disease risk factors (blood lipids from 188 577 subjects, type 2 diabetes from 148 821 subjects, body mass index from 123 865 subjects, and smoking phenotypes from 74 053 subjects). We found that 6p22.1 (rs6904596, ZNF184) was associated with both lung cancer ($P = 5.50 \times 10^{-6}$) and blood triglycerides ($P = 1.39 \times 10^{-5}$). We replicated the association in 6097 lung cancer case patients and 204 657 control subjects ($P = 2.40 \times 10^{-4}$) and in 71 113 subjects with triglycerides data ($P = .01$). rs6904596 reached genome-wide significance in lung cancer meta-analysis (odds ratio = 1.15, 95% confidence interval = 1.10 to 1.21, $P_{combined} = 5.20 \times 10^{-9}$). The large sample size provided by the lipid GWAS data and the shared genetic risk factors between the two traits contributed to the uncovering of a hitherto unidentified genetic locus for lung cancer.

Genetic heritability of lung cancer is estimated to be 14% [1], but only a few genetic risk loci have been identified to date in genome-wide association studies (GWAS) of lung cancer in Europeans [2]. Epidemiological studies have shown associations between lung cancer and cardiovascular disease (CVD) risk factors related to the metabolic syndrome [3,4]. There is also substantial evidence that lipid metabolism and innate immunity evolved from common pathways, and consequently genes that influence lipid traits may also influence inflammation and subsequent cancer development [5–7]. Lung cancer is also well known to be strongly associated with tobacco smoking. Predicated on the hypothesis that investigating shared genetic risk factors across these traits could enhance the possibility of identifying new genetic loci for lung cancer, we used quantile-quantile (Q-Q) plots [8] (Supplementary Methods, available online) to assess potential polygenic enrichment of single-nucleotide polymorphisms (SNPs) associated with lung cancer given association with each CVD risk factor or smoking phenotypes (Figure 1; Supplementary Figure 1, available online).

The analysis was based on the TRICL consortium meta-analysis of lung cancer GWAS, including 12 160 lung cancer case patients and 16 838 control subjects [2] (Supplementary Table 1, available online); the meta-analysis data of blood lipids from the Global Lipids Genetics Consortium (GLGC; including genetic association with triglycerides [TG], and high and low density lipoproteins-cholesterol [HDL-C and LDL-C]) from 188 577 subjects [9], of type 2 diabetes (T2D) from 148 821 subjects [10], and of body mass index (BMI) from 123 865 subjects [11]; and the meta-analysis of cigarettes per day (CPD) and never vs ever smoking (SMOKER) data from the Tobacco, Alcohol and Genetics (TAG) consortium, including 74 053 subjects (Supplementary Table 2, available online). The Supplementary Materials (available online) contain additional details on the contributing studies, statistical analyses, and functional tests. All statistical tests were two-sided unless otherwise specified. P values of less than .05 were considered statistically significant.

The Q-Q plots show enrichment between lung cancer and LDL-C and between lung cancer and TG blood lipid traits across multiple P value thresholds, up to 10^{-5} (Figure 1, A and B), verified by an adaptive permutation procedure (Supplementary Table 3, available online). In contrast, we observed no statistically significant enrichment (Bonferroni-corrected significance threshold $P < .0036$) between lung cancer and HDL-C, BMI, T2D, or smoking phenotypes (the analysis of smoking excluded the SNP markers mapping to chr15:78,686,690-79,231,478, which are known to be associated with lung cancer and smoking (Supplementary Table 3 and Supplementary Figure 1, available online) [12–13]. Thus, we excluded these traits from further analysis.
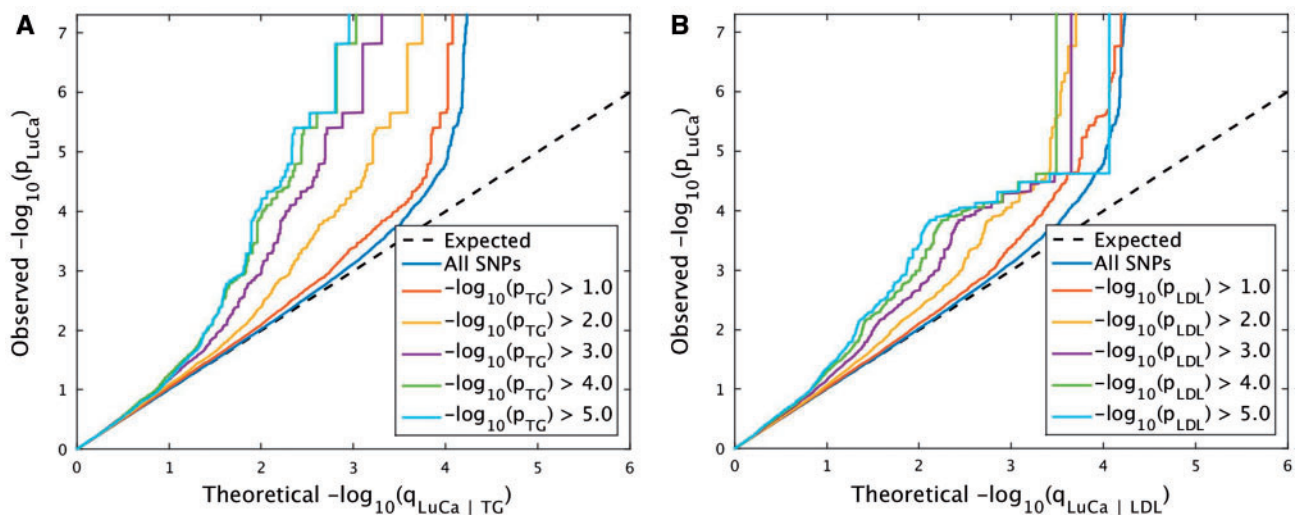

Figure 1. Conditional Q-Q plots: lung cancer (LuCa) | cardiovascular disease (CVD) factors (triglycerides [TG] and low-density lipoproteins-cholesterol [LDL-C]). 'Conditional Q-Q plot' of theoretical vs empirical $-\log_{10}$ P values (corrected for genomic control λ) in LuCa below the standard GWAS threshold of $-\log_{10}$ P values equal to 7.3 (equals P values above 5×10^{-8}) as a function of statistical significance of association with **(A)** TG and **(B)** LDL-C at the levels of $P < 1.00$, $P < .10$, $P < .01$, $P < .001$, $P < 1.00 \times 10^{-4}$, $P < 1.00 \times 10^{-5}$, respectively. **Dotted lines** indicate the theoretical line in case of no association. All statistical tests were two-sided.

Cross-phenotype-associated loci between lung cancer and TG and between lung cancer and LDL-C were assessed by conjunction false discovery rate (FDR) (Supplementary Materials, available online) (8). Because controlling FDR is heavily affected by the number of identified SNPs, we pruned SNPs in linkage disequilibrium (LD; $r^2 > 0.8$) and excluded the major histocompatibility complex (MHC; genomic position (hg 19): chr6:29,528,318-33,373,649 [14]), which harbors established lung cancer susceptibility SNPs and is known for long-range LD. By controlling conjunction FDR, we identified one genetic locus at 6p22.1, rs6904596 (A>G, minor allele frequency in caucasians = 0.094), associated with both lung cancer and blood triglycerides (conjunction FDR = 1.24E-02; $P = 5.50 \times 10^{-6}$ for lung cancer; $P = 1.39 \times 10^{-5}$ for TG) (this locus and additional genetic loci shared between lung cancer and lipid traits are shown in Supplementary Table 4 and Supplementary Figures 2-6, available online). This locus remained statistically significant as defined by a conjunction FDR of less than .05, also using different thresholds for pruning SNPs in LD (Supplementary Table 5, available online).

We tested this SNP for replication in 6097 lung cancer case patients and 204 657 control subjects from deCODE, Harvard, Holland, and Spain (Supplementary Table 6, available online). This locus was replicated ($P_{replication} = 2.40 \times 10^{-4}$) and attained genome-wide significance for lung cancer risk in the meta-analysis of discovery and replication data (odds ratio = 1.15, 95% confidence interval = 1.10 to 1.21, two-sided $P_{combined} = 5.20 \times 10^{-9}$, $P_{heterogeneity} = .91$) (Table 1). This SNP was also replicated in the association with TG in 71 113 independent samples from deCODE and Holland (two-sided $P = .01$, $P_{combined} = 1.34 \times 10^{-6}$) (Table 1). The SNP association with lung cancer was mostly driven by the squamous cell carcinoma subtype ($P = 2.80 \times 10^{-5}$) and not adenocarcinoma ($P = 0.06$) (Supplementary Table 7, available online).

rs6904596 localizes to 6p22.1 (27,491,299 bp; hg19) and lies 50 kb 5' of zinc finger protein 184 (ZNF184). It shows expression-QTL in lung tissue (15) with HLA-DRB3 ($\beta = -6.79$, $P = 1.10 \times 10^{-11}$). Additionally, rs7749305, located on chr6:27,446,566 ($r^2 = 1$ with rs6904596 in HapMap 3 of Caucasian populations), shows suggestive regulatory functions. This SNP showed the strongest association with lung cancer but was not genotyped in the Global Lipids Genetics Consortium GWAS. It lies within a DNaseI hypersensitive region in small airway epithelial cells (SAEC) and A549 adenocarcinoma cells (ENCODE) and lies in a region hypomethylated in primary alveolar epithelial cells (AEC) from our laboratory (Supplementary Figure 7, available online). rs7749305 alternate allele C appears to create ATF3 and HIF1A binding

sites. Similar findings are evident in adipocytes (ENCODE), extending the pleiotropic association between lung cancer and lipid traits to their function in respective tissue types.

Additional functional analyses on relevant tissue types would be important to study how the genetic risk variants assert their downstream effect on the traits of interest.

Our study emphasizes that pleiotropic analysis of GWAS data of epidemiologically related traits can uncover hitherto unidentified genetic associations. Moreover, some GWAS of quantitative traits may be much larger than disease-specific GWAS (like in the case of CVD risk factors vs lung cancer) and, thus, may improve the likelihood to identify new loci for the disease with the smaller sample size.

Funding

Table 1. Association of rs6904596 at 6p22.1 with both lung cancer risk and blood triglyceride levels

Study	Lung cancer cases/ controls	OR (95% CI)	P*	Study	Triglyceride individuals	β (95% CI)	P*
TRICL	12160/16838	1.15 (1.08 to 1.21)	5.50×10^{-6}	GLGC	188577	0.0244 (0.013 to 0.035)	1.39×10^{-5}
Replication	6097/204657	1.16 (1.07 to 1.25)	2.40×10^{-4}	Replication	71113	0.0290 (0.006 to 0.051)	1.14×10^{-2}
deCODE	3865/196658	1.17 (1.06 to 1.30)	3.05×10^{-3}	deCODE	66027	0.0200 (-0.004 to 0.044)	.10
Harvard	984/ 970	1.18 (0.93 to 1.50)	.17	Holland	5086	0.0891 (0.027 to 0.151)	5.12×10^{-3}
Holland	687/ 5158	1.15 (0.96 to 1.37)	.12				
Spain	561/ 1871	1.10 (0.88 to 1.37)	.40	Combined	259690	0.0253 (0.015 to 0.035)	1.34×10^{-6}
Combined	18257/ 221495	1.15 (1.10 to 1.21)	5.20×10^{-9}				
P_Q-heterogeneity†			.91				.23

*P values were derived from a two-sided Wald test. The reference group for the odds ratio in the lung cancer study was healthy controls without lung cancer. CI = confidence interval; OR = odds ratio.

†Heterogeneity of effect size across studies was evaluated using the Cochran's Q statistic. The test is defined as one-sided.

Notes

References

1. Hemminki K, Lonnstedt I, Vaittinen P, et al. Estimation of genetic and environmental components in colorectal and lung cancer and melanoma. *Genet Epidemiol*. 2001;20(1):107–116.
2. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*. 2014;46(7):736–741.
3. Braun S, Bitton-Worms K, LeRoith D. The link between the metabolic syndrome and cancer. *Int J Biol Sci*. 2011;7(7):1003–1015.
4. Kucharska-Newton AM, Rosamond WD, Schroeder JC, et al. HDL-cholesterol and the incidence of lung cancer in the Atherosclerosis Risk in Communities (ARIC) study. *Lung Cancer*. 2008;61(3):292–300.
5. Kominsky DJ, Campbell EL, Colgan SP. Metabolic shifts in immunity and inflammation. *J Immunol*. 2010;184(8):4062–4068.
6. Shi J, Chatterjee N, Rotunno M, et al. Inherited variation at chromosome 12p13.33, including RAD52, influences the risk of squamous cell lung carcinoma. *Cancer Discov*. 2012;2(2):131–139.
7. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. *Nat Rev Cancer*. 2009;9(11):798–809.
8. Andreassen OA, Zuber V, Thompson WK, et al. Shared common variants in prostate cancer and blood lipids. *Int J Epidemiol*. 2014;43(4):1205–1214.
9. Global Lipids Genetics C, Willer CJ, Schmidt EM, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274–1283.
10. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. 2012;44(9):981–990.
11. Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010; 42(11):937–948.
12. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008; 40(5):616–622.
13. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008; 452(7187):638–642.
14. Shiina T, Hosomichi K, Inoko H, et al. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Genet*. 2009;54(1):15–39.
15. Hao K, Bosse Y, Nickle DC, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet*. 2012;8(11):e1003029.