



**QUEEN'S
UNIVERSITY
BELFAST**

Pleiotropic Analysis of Lung Cancer and Blood Triglycerides

Zuber, V., Marconett, C. N., Shi, J., Hua, X., Wheeler, W., Yang, C., Song, L., Dale, A. M., Laplana, M., Risch, A., Witoelar, A., Thompson, W. K., Schork, A. J., Bettella, F., Wang, Y., Djurovic, S., Zhou, B., Borok, Z., van der Heijden, H. F. M., ... Landi, M. T. (2016). Pleiotropic Analysis of Lung Cancer and Blood Triglycerides. *Journal of the National Cancer Institute*, 108(12). <https://doi.org/10.1093/jnci/djw167>

Published in:

Journal of the National Cancer Institute

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

Copyright © 2016, Oxford University Press

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in Journal of the National Cancer Institute following peer review. The definitive publisher-authenticated version Zuber, V, Marconett, CN, Shi, J, Hua, X, Wheeler, W, Yang, C, Song, L, Dale, AM, Laplana, M, Risch, A, Witoelar, A, Thompson, WK, Schork, AJ, Bettella, F, Wang, Y, Djurovic, S, Zhou, B, Borok, Z, van der Heijden, HFM, de Graaf, J, Swinkels, D, Aben, KK, McKay, J, Hung, RJ, Bikeböller, H, Stevens, VL, Albanes, D, Caporaso, NE, Han, Y, Wei, Y, Panadero, MA, Mayordomo, JI, Christiani, DC, Kiemeny, L, Andreassen, OA, Houlston, R, Amos, CI, Chatterjee, N, Laird-Offringa, IA, Mills, IG & Landi, MT 2016, 'Pleiotropic Analysis of Lung Cancer and Blood Triglycerides' Journal of the National Cancer Institute, vol 108, no. 12. is available online at: <http://jnci.oxfordjournals.org/content/108/12/djw167>

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Pleiotropic analysis of lung cancer and blood triglycerides identifies a shared genetic locus

Verena Zuber^{1,2,3,4}; Crystal N. Marconett⁵; Jianxin Shi⁶; Xing Hua⁶; William Wheeler⁷; Chenchen Yang⁵; Lei Song⁶; Anders M. Dale^{8,9,10,11}; Marina Laplana¹²; Angela Risch^{12,13,14}; Aree Witoelar^{1,2}; Wesley K. Thompson¹⁵; Andrew J. Schork^{8,9,16}; Francesco Bettella^{1,2}; Yunpeng Wang^{1,2}; Srdjan Djurovic^{17,18}; Beiyun Zhou¹⁹; Zea Borok¹⁹; Henricus F.M. van der Heijden²⁰; Jacqueline de Graaf²⁰; Dorine Swinkels²¹; Katja K. Aben²²; James McKay²³; Rayjean J. Hung²⁴; Heike Bikeböllner²⁵; Victoria L. Stevens²⁶; Demetrius Albanes⁶; Neil E. Caporaso⁶; Younghun Han²⁷; Yongyue Wei²⁸; Maria Angeles Panadero²⁸; Jose I Mayordomo²⁹; David C. Christiani^{28,31}; Lambertus Kiemeny²⁰; Ole A. Andreassen^{1,2}; Richard Houlston³²; Christopher I. Amos²⁷; Nilanjjan Chatterjee⁶; Ite A. Laird-Offringa⁵; Ian G. Mills^{3,33,34†}; Maria Teresa Landi^{6†}

† These authors contributed equally

Correspondent authors: Maria Teresa Landi (landim@mail.nih.gov) and Ian Mills (I.Mills@qub.ac.uk; ian.mills@ncmm.uio.no)

¹NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

²Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

³Prostate Cancer Research Group, Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo and Oslo University Hospital, Oslo, Norway

⁴European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK.

⁵Departments of Surgery and of Biochemistry and Molecular Biology, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

⁶Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, U.S. Public Health Service, Bethesda, MD 20892, USA

⁷Information Management Services, Inc.; Rockville, MD, 20852; USA

⁸Multimodal Imaging Laboratory, University of California at San Diego, La Jolla, CA, USA

⁹Center for Human Development, University of California at San Diego, La Jolla, CA, USA

¹⁰Department of Radiology, University of California, San Diego, La Jolla, CA, USA

¹¹Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

¹²Division of Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany.

¹³Department of Molecular Biology, University of Salzburg, Salzburg, Austria.

¹⁴Translational Lung Research Center Heidelberg TLRC-H, Member of the German Center for Lung Research DZL, Heidelberg, Germany.

¹⁵Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

¹⁶Cognitive Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA

¹⁷Department of Medical Genetics, Oslo University Hospital, Oslo, Norway

¹⁸NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway

¹⁹Will Rogers Institute Pulmonary Research Center and Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, and Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

²⁰Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands

²¹Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

- ²²Netherlands Comprehensive Cancer Organization, Utrecht, The Netherlands
- ²³International Agency for Research on Cancer (IARC, World Health Organization (WHO)), Lyon, France.
- ²⁴Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada.
- ²⁵Zentrum Informatik, Statistik und Epidemiologie, Universitätsmedizin Göttingen
- ²⁶Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA.
- ²⁷Center for Genomic Medicine, Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, New Hampshire, USA.
- ²⁸Department of Epidemiology and Environmental Health, Harvard School of Public Health
- ²⁹Division of Medical Oncology, Ciudad de Coria Hospital, Coria, Spain
- ³⁰Division of Medical Oncology, University Hospital, Zaragoza, Spain
- ³¹Division of Pulmonary/Critical Care, Department of Medicine Massachusetts General Hospital/Harvard Medical School
- ³²Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK.
- ³³Department of Molecular Oncology, Institute of Cancer Research and Department of Urology, Oslo University Hospital, Oslo, Norway
- ³⁴Prostate Cancer UK/Movember Centre of Excellence for Prostate Cancer Research, Centre for Cancer Research and Cell Biology, Queen's University, Belfast, UK

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 cases and 16,838 controls) 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 ($P=5.5 \times 10^{-6}$) and blood triglycerides ($P=1.39 \times 10^{-5}$). We replicated the association in 6,097 lung cancer cases and 204,657 controls ($P=2.4 \times 10^{-4}$) and in 71,113 subjects with triglycerides data ($P=0.011$). rs6904596 reached genome-wide significance in lung cancer meta-analysis (odds ratio=1.15, $P_{\text{combined}}=5.2 \times 10^{-9}$). 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 to identify new genetic loci for lung cancer, we used quantile-quantile (Q-Q) plots [8] (Online Methods) to assess potential polygenic enrichment of SNPs associated with lung cancer given association with each CVD risk factor or smoking phenotypes (**Figure 1 and Supplementary Figure 1**).

The analysis was based on the TRICL consortium meta-analysis of lung cancer GWAS, including 12,160 lung cancer cases and 16,838 controls [2] (**Supplementary Table 1**); 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 data from the Tobacco, Alcohol and Genetics (TAG) consortium, including 74,053 subjects (**Supplementary Table 2**). The **Supplementary Materials** (available online) contain additional details on the contributing studies, statistical analyses and functional tests.

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 1A-B**) verified by an adaptive permutation procedure (**Supplementary Table 3**). In contrast, we observed no significant enrichment ($P < 0.001$) between lung cancer and HDL, 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 [12-13] (**Supplementary Table 3**,

Supplementary Figure 1). Thus we excluded these traits from further analysis.

Cross-phenotype associated loci between lung cancer and TG and between lung cancer and LDL-C were assessed by conjunction false discovery rate (FDR) [8] (**Supplementary Materials**). 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=0.0124; $P=5.5 \times 10^{-6}$ 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**). This locus remained significant also using different thresholds for pruning SNPs in LD (**Supplementary Table 5**).

We tested this SNP for replication in 6,097 lung cancer cases and 204,657 controls from deCODE, Harvard, Holland and Spain (**Supplementary Table 6**). This locus was replicated ($P_{\text{replication}}=2.4 \times 10^{-4}$) and attained genome-wide significance for lung cancer risk in the meta-analysis of discovery and replication data (two-sided $P_{\text{combined}}=5.2 \times 10^{-9}$, $P_{\text{heterogeneity}}=0.91$, **Table 1**). This SNP was also replicated in the association with TG in 71,113 independent samples from deCODE and Holland (two-sided $P=0.011$, $P_{\text{combined}}=1.34 \times 10^{-6}$, **Table 1**).

The SNP association with lung cancer was mostly driven by the squamous cell carcinoma subtype ($P=2.8 \times 10^{-5}$) and not adenocarcinoma ($P=0.06$, **Supplementary**

Table 7).

rs6904596 localizes to 6p22.1 (27,491,299 bp; hg19) and lies 50kb 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 Lipid 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**). 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.

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

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD.

Transdisciplinary Research for Cancer of Lung (TRICL): National Institute of Health U19 CA148127-01 (PI: Amos), Canadian Cancer Society Research Institute (no. 020214, PI: Hung). The Environment and Genetics in Lung Cancer Etiology (EAGLE), Prostate, Lung, Colon, Ovary Screening Trial (PLCO), and Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) studies were supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute (NCI), Division of Cancer Epidemiology and Genetics. ATBC was also supported by U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, and N01-RC-37004) from the NCI.

PLCO was also supported by individual contracts from the NCI to the University of Colorado Denver (NO1-CN-25514), Georgetown University (NO1-CN-25522), the Pacific Health Research Institute (NO1-CN-25515), the Henry Ford Health System (NO1-CN-25512), the University of Minnesota, (NO1-CN-25513), Washington University (NO1-CN-25516), the University of Pittsburgh (NO1-CN-25511), the University of Utah (NO1-CN-25524), the Marshfield Clinic Research Foundation (NO1-CN-25518), the University of Alabama at Birmingham (NO1-CN-75022), Westat, Inc. (NO1-CN-25476), and the University of California, Los Angeles (NO1-CN-25404). The Cancer Prevention Study-II (CPS-II) Nutrition Cohort was supported by the American Cancer Society. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping. Harvard Lung Study: The Harvard Lung Cancer Study is supported by the US National Institutes of Health (R01 CA092824, P50 CA090578, and R01 CA074386).

The authors thank deCODE genetics for contributing GWAS data.

Global Lipids Genetics Consortium (GLGC): Data on the lipid traits were provided by the Global Lipids Consortium through their access portal (<http://csg.sph.umich.edu/abecasis/public/lipids2013/>). The full Consortium acknowledgements are available in the supplementary information of Willer *et al.*

Tobacco, Alcohol and Genetics (TAG) consortium: Data on the smoking traits were provided by the Tobacco, Alcohol and Genetics through their access portal (<http://www.broadinstitute.org/mpg/ricopili/>). The full Consortium acknowledgements are available in the supplementary information of *the Tobacco, Alcohol and Genetics*.

Lung eQTL study: The lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research

(MOP - 123369), and the Cancer Research Society and Read for the Cure. Y. Bossé is the recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec – Santé (FRQS).

Lung meQTL study: The meQTL study based on the environment and Genetics in Lung Cancer Etiology (EAGLE) study was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS.

Epigenetics: This study was supported by NIH grant (1 R01 HL114094) to IAL-O and ZB, NIH grants (1 P30 H101258) and (R37HL062569-13) to ZB. ZB is supported by the Ralph Edgington Chair in Medicine. CNM was supported in part by the department of Surgery, USC. Generation of epigenetic data was supported in part by the Norris Comprehensive Cancer Center core grant, award number P30CA014089 from the National Cancer Institute.

This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

Notes

The study funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; nor the decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The authors have no conflicts of interest to disclose.

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):736-41.
3. Braun S, Bitton-Worms K, LeRoith D. The link between the metabolic syndrome and cancer. *Int J Biol Sci* 2011;7(7):1003-15.
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):4062-8.
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):131-9.
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; 10.1093/ije/dyu090.
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):1274-83.

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):981-90.
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):937-48.
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):616-22.
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):638-42.
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.

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

Study	Lung cancer Cases/ Controls	OR	95% CI	P-value		Study	Triglycerides Individuals	β	95% CI	P-value
TRICL	12160/16838	1.15	(1.08,1.21)	5.50×10^{-6}		GLGC	188577	0.0244	(0.013,0.035)	1.39×10^{-5}
Replication	6097/204657	1.16	(1.07,1.25)	2.4×10^{-4}		Replication	71113	0.0290	(0.006,0.051)	1.14×10^{-2}
deCODE	3865/196658	1.17	(1.06,1.30)	3.05×10^{-3}		deCODE	66027	0.0200	(-0.004,0.044)	0.10
Harvard	984/ 970	1.18	(0.93,1.50)	0.171		Holland	5086	0.0891	(0.027,0.151)	5.12×10^{-3}
Holland	687/ 5158	1.15	(0.96,1.37)	0.119						
Spain	561/ 1871	1.10	(0.88,1.37)	0.40		Combined	259690	0.0253	(0.015,0.035)	1.34×10^{-6}
Combined P _Q - heterogeneity*	18257/ 221495	1.15	(1.10,1.21)	5.2×10^{-9} 0.92						0.23

* Heterogeneity of effect size across studies was evaluated using the Cochran's Q statistic

Figure legends

Figure 1. Conditional Q-Q plots: LuCa | CVD factors (TG and LDL-C).

‘Conditional Q-Q plot’ of theoretical vs empirical $-\log_{10} p$ -values (corrected for genomic control λ) in lung cancer (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 significance of association with (A) triglycerides (TG) and (B) low-density lipoprotein (LDL-C) at the level of $p < 1$, $p < 0.1$, $p < 0.01$, $p < 0.001$, $p < 0.0001$, $p < 0.00001$ respectively. Dotted lines indicate the theoretical line in case of no association.