

# Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function

Dana B. Hancock<sup>1,2,9</sup>, María Soler Artigas<sup>3,9</sup>, Sina A. Gharib<sup>4,5,9</sup>, Amanda Henry<sup>6,9</sup>, Ani Manichaikul<sup>7,8,9</sup>, Adaikalavan Ramasamy<sup>9,10,11,9</sup>, Daan W. Loth<sup>12,13,9</sup>, Medea Imboden<sup>14,15</sup>, Beate Koch<sup>16</sup>, Wendy L. McArdle<sup>17</sup>, Albert V. Smith<sup>18,19</sup>, Joanna Smolonska<sup>20</sup>, Akshay Sood<sup>21</sup>, Wenbo Tang<sup>22</sup>, Jemma B. Wilk<sup>23,24</sup>, Guangju Zhai<sup>25,26</sup>, Jing Hua Zhao<sup>27</sup>, Hugues Aschard<sup>28</sup>, Kristin M. Burkart<sup>29</sup>, Ivan Curjuric<sup>14,15</sup>, Mark Eijgelsheim<sup>12</sup>, Paul Elliott<sup>10,30</sup>, Xiangjun Gu<sup>31</sup>, Tamara B. Harris<sup>32</sup>, Christer Janson<sup>33</sup>, Georg Homuth<sup>34</sup>, Pirro G. Hysi<sup>25</sup>, Jason Z. Liu<sup>35</sup>, Laura R. Loehr<sup>36</sup>, Kurt Lohman<sup>37</sup>, Ruth J. F. Loos<sup>27</sup>, Alisa K. Manning<sup>38,39,40</sup>, Kristin D. Marcianti<sup>5</sup>, Ma'en Obeidat<sup>6</sup>, Dirkje S. Postma<sup>41,42</sup>, Melinda C. Aldrich<sup>43</sup>, Guy G. Brusselle<sup>44</sup>, Ting-hsu Chen<sup>45,46</sup>, Gudny Eiriksdottir<sup>18</sup>, Nora Franceschini<sup>36</sup>, Joachim Heinrich<sup>47</sup>, Jerome I. Rotter<sup>48</sup>, Cisca Wijmenga<sup>20</sup>, O. Dale Williams<sup>49</sup>, Amy R. Bentley<sup>50</sup>, Albert Hofman<sup>12</sup>, Cathy C. Laurie<sup>51</sup>, Thomas Lumley<sup>52</sup>, Alanna C. Morrison<sup>53</sup>, Bonnie R. Joubert<sup>2</sup>, Fernando Rivadeneira<sup>12,54,55</sup>, David J. Couper<sup>56</sup>, Stephen B. Kritchevsky<sup>57</sup>, Yongmei Liu<sup>58</sup>, Matthias Wjst<sup>59,60</sup>, Louise V. Wain<sup>3</sup>, Judith M. Vonk<sup>42,61,62</sup>, André G. Uitterlinden<sup>12,54,55</sup>, Thierry Rochat<sup>63</sup>, Stephen S. Rich<sup>7</sup>, Bruce M. Psaty<sup>64,65,66</sup>, George T. O'Connor<sup>24,46</sup>, Kari E. North<sup>36</sup>, Daniel B. Mirel<sup>67</sup>, Bernd Meibohm<sup>68</sup>, Lenore J. Launer<sup>32</sup>, Kay-Tee Khaw<sup>69</sup>, Anna-Liisa Hartikainen<sup>70</sup>, Christopher J. Hammond<sup>25</sup>, Sven Gläser<sup>16</sup>, Jonathan Marchini<sup>35</sup>, Peter Kraft<sup>71</sup>, Nicholas J. Wareham<sup>27</sup>, Henry Völzke<sup>72</sup>, Bruno H. C. Stricker<sup>12,13,54,73</sup>, Timothy D. Spector<sup>25</sup>, Nicole M. Probst-Hensch<sup>14,15</sup>, Deborah Jarvis<sup>9,30</sup>, Marjo-Riitta Jarvelin<sup>10,30,74,75,76</sup>, Susan R. Heckbert<sup>64,66</sup>, Vilundur Gudnason<sup>18,19</sup>, H. Marika Boezen<sup>42,61,62</sup>, R. Graham Barr<sup>29,77</sup>, Patricia A. Cassano<sup>22,78¶</sup>, David P. Strachan<sup>79¶</sup>, Myriam Fornage<sup>31,53¶</sup>, Ian P. Hall<sup>6¶</sup>, José Dupuis<sup>24,80¶</sup>, Martin D. Tobin<sup>3¶\*</sup>, Stephanie J. London<sup>2,81¶\*</sup>

**1** Behavioral Health Epidemiology Program, Research Triangle Institute International, Research Triangle Park, North Carolina, United States of America, **2** Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, Research Triangle Park, North Carolina, United States of America, **3** Departments of Health Sciences and Genetics, University of Leicester, Leicester, United Kingdom, **4** Center for Lung Biology, University of Washington, Seattle, Washington, United States of America, **5** Department of Medicine, University of Washington, Seattle, Washington, United States of America, **6** Division of Therapeutics and Molecular Medicine, University of Nottingham, Queen's Medical Centre, Nottingham, United Kingdom, **7** Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, United States of America, **8** Department of Public Health Sciences, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, Virginia, United States of America, **9** Respiratory Epidemiology and Public Health, Imperial College London, London, United Kingdom, **10** Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom, **11** Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, United Kingdom, **12** Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, **13** Inspectorate of Healthcare, The Hague, The Netherlands, **14** Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland, **15** University of Basel, Basel, Switzerland, **16** Department of Internal Medicine B, University Hospital Greifswald, Greifswald, Germany, **17** ALSPAC, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, **18** Icelandic Heart Association, Kopavogur, Iceland, **19** University of Iceland, Reykjavik, Iceland, **20** Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, **21** Department of Medicine, University of New Mexico, Albuquerque, New Mexico, United States of America, **22** Division of Nutritional Sciences, Cornell University, Ithaca, New York, United States of America, **23** Division of Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, **24** The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **25** Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **26** Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada, **27** MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom, **28** Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **29** Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, United States of America, **30** MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom, **31** Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **32** National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, **33** Department of Medical Sciences, Respiratory Medicine, Uppsala University, Uppsala, Sweden, **34** Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University of Greifswald, Greifswald, Germany, **35** Department of Statistics, University of Oxford, United Kingdom, **36** Gillings School of Global Public Health, Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **37** Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **38** Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **39** Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **40** Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, **41** Department of Pulmonology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, **42** GRIAC Research Institute, University Medical Center Groningen, Groningen, The Netherlands, **43** Department of Thoracic Surgery and Division of Epidemiology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **44** Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium, **45** Section of Pulmonary and Critical Care Medicine, Department of Medicine, Veterans Administration Boston Healthcare System, Boston, Massachusetts, United States of America, **46** The Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **47** Institute of Epidemiology I, Helmholtz Zentrum München, Munich, Germany, **48** Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles,

California, United States of America, **49** Department of Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, Florida, United States of America, **50** Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **51** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, **52** Department of Statistics, University of Auckland, Auckland, New Zealand, **53** Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, United States of America, **54** Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands, **55** Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, The Netherlands, **56** Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **57** Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **58** Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **59** Institute of Lung Biology and Disease, Comprehensive Pneumology Center, Helmholtz Zentrum München, Neuherberg, Germany, **60** Institute for Medical Statistics and Epidemiology (IMSE), Technical University Munich, Munich, Germany, **61** Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands, **62** The LifeLines Cohort Study, Groningen, The Netherlands, **63** Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland, **64** Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, Washington, United States of America, **65** Departments of Medicine and Health Services, University of Washington, Seattle, United States of America, **66** Group Health Research Institute, Group Health Cooperative, Seattle, Washington, **67** Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **68** College of Pharmacy, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **69** Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, **70** Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Oulu, Oulu, Finland, **71** Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, **72** Institute for Community Medicine, Study of Health In Pomerania (SHIP)/Clinical Epidemiological Research, University of Greifswald, Greifswald, Germany, **73** Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands, **74** Department of Children, Young People, and Families, National Institute for Health and Welfare, Oulu, Finland, **75** Institute of Health Sciences, University of Oulu, Oulu, Finland, **76** Biocenter Oulu, University of Oulu, Oulu, Finland, **77** Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, United States of America, **78** Department of Public Health, Weill Cornell Medical College, New York, New York, United States of America, **79** Division of Population Health Sciences and Education, St. George's University of London, London, United Kingdom, **80** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **81** Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, U.S. Department of Health and Human Services, Research Triangle Park, North Carolina, United States of America

## Abstract

Genome-wide association studies have identified numerous genetic loci for spirometric measures of pulmonary function, forced expiratory volume in one second ( $FEV_1$ ), and its ratio to forced vital capacity ( $FEV_1/FVC$ ). Given that cigarette smoking adversely affects pulmonary function, we conducted genome-wide joint meta-analyses (JMA) of single nucleotide polymorphism (SNP) and SNP-by-smoking (ever-smoking or pack-years) associations on  $FEV_1$  and  $FEV_1/FVC$  across 19 studies (total  $N = 50,047$ ). We identified three novel loci not previously associated with pulmonary function. SNPs in or near *DNER* (smallest  $P_{JMA} = 5.00 \times 10^{-11}$ ), *HLA-DQB1* and *HLA-DQA2* (smallest  $P_{JMA} = 4.35 \times 10^{-9}$ ), and *KCNJ2* and *SOX9* (smallest  $P_{JMA} = 1.28 \times 10^{-8}$ ) were associated with  $FEV_1/FVC$  or  $FEV_1$  in meta-analysis models including SNP main effects, smoking main effects, and SNP-by-smoking (ever-smoking or pack-years) interaction. The HLA region has been widely implicated for autoimmune and lung phenotypes, unlike the other novel loci, which have not been widely implicated. We evaluated *DNER*, *KCNJ2*, and *SOX9* and found them to be expressed in human lung tissue. *DNER* and *SOX9* further showed evidence of differential expression in human airway epithelium in smokers compared to non-smokers. Our findings demonstrated that joint testing of SNP and SNP-by-environment interaction identified novel loci associated with complex traits that are missed when considering only the genetic main effects.

**Citation:** Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, et al. (2012) Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. *PLoS Genet* 8(12): e1003098. doi:10.1371/journal.pgen.1003098

**Editor:** Greg Gibson, Georgia Institute of Technology, United States of America

**Received:** July 3, 2012; **Accepted:** October 1, 2012; **Published:** December 20, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported, in part, by the Intramural Research Program of the National Institutes of Health (NIH), the National Institute of Environmental Health Sciences (NIEHS, Z01E5043012). The CHARGE Pulmonary Working Group acknowledges funding from the National Heart, Lung, and Blood Institute (NHLBI) (HL105756) and organizational support from the CHARGE Consortium. From SpiroMeta, MD Tobin was supported by UK MRC Senior Clinical Fellowship G0902313; IP Hall and the laboratory work on expression profiling were supported by MRC (G1000861). P Kraft and H Aschard were supported by R21DK084529. The Age, Gene/Environment Susceptibility (AGES)-Reykjavik Study is funded by NIH contract number N01-AG-12100, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute (NHGRI) contract U01HG004402; and NIH contract HHSN268200625226C. Infrastructure was partly supported by grant number UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research. The British 1958 Cohort (B58C) DNA collection was funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. Genotyping for the Wellcome Trust Case Control Consortium of B58C was funded by the Wellcome Trust grant 076113/B/04/Z. The Type 1 Diabetes Genetics Consortium of B58C was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, NHGRI, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International and supported by U01 DK062418. Genome-wide data was deposited by the Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile Diabetes Research Foundation International, the Wellcome Trust, and the National Institute for Health Research Cambridge Biomedical Research Centre and is in receipt of a Wellcome Trust Strategic Award (079895). The Coronary Artery Risk Development in Young Adults (CARDIA) study was funded by contracts N01-HC-95095, N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, N01-HC-45134, N01-HC-05187, N01-HC-45205, and N01-HC-45204 from NHLBI to the CARDIA investigators. Genotyping of the CARDIA participants was supported by grants U01-HG-004729, U01-HG-004446, and U01-HG-004424 from the NHGRI. Statistical analyses were supported by grants U01-HG-004729 and R01-HL-084099 to M Fornage Cardiovascular Health Study (CHS) was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, and HHSN268201200036C and by NHLBI grants HL080295, HL075366, HL087652, and HL105756 with additional contribution from

National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the National Institute on Aging (NIA) and the Cedars-Sinai Board of Governors' Chair in Medical Genetics (Jl Rotter). DNA handling and genotyping was supported in part by National Center for Research Resources CTSI grant UL 1RR033176 and NIDDK grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The European Community Respiratory Health Survey (ECRHS) acknowledges funding from the European Union (GABRIEL GRANT Number: 018996, ECRHS II Coordination Number: QLK4-CT-1999-01237). The European Prospective Investigation of Cancer (EPIC)-Norfolk Study is funded by Cancer Research UK and the Medical Research Council. Framingham Heart Study (FHS) research was conducted in part using data and resources of the NHLBI and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the FHS investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by NHLBI (contract no. N01-HC-25195) and its contract with Affymetrix for genotyping services (contract no. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. JB Wilk was supported by a Young Clinical Scientist Award from the Flight Attendant Medical Research Institute (FAMRI). Health, Aging, and Body Composition (Health ABC) was supported by NIA contracts N01AG62101, N01AG2103, and N01AG62106, and in part by the Intramural Research Program of NIA. This work was also supported, in part, by Intramural Research Programs of the NHGRI. The genome-wide association study (GWAS) in Health ABC was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences, and genotyping services were provided by the Center for Inherited Disease Research, which is fully funded through an NIH contract to The Johns Hopkins University (HHSN268200782096C). This research was further supported by RC1AG035835. The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study, is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006); the Economic Structure Enhancing Fund (FES) of the Dutch government; the Ministry of Economic Affairs; the Ministry of Education, Culture, and Science; the Ministry for Health, Welfare, and Sports; the Northern Netherlands Collaboration of Provinces (SNN); the Province of Groningen; University Medical Center Groningen; the University of Groningen; Dutch Kidney Foundation; and Dutch Diabetes Research Foundation. We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergero, and Rosalie Visser for their work related to data-collection and validation. The Multi-Ethnic Study of Atherosclerosis (MESA) study was supported by contracts N01-HC-95159 through N01-HC-95169 from the NHLBI and RR-024156. The MESA Lung study was supported by grants R01-HL077612 and RC1-HL100543 from the NHLBI. Funding for SHARe genotyping was provided by NHLBI contract N02-HL-6-4278. The Northern Finland Birth Cohort of 1966 (NFBC1966) (to M-R Jarvelin) received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/National Institute of Mental Health (5R01MH63706-02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (PrevMetSyn/SALVE). P Elliott is a National Institute of Health Research (NIHR) Senior Investigator and acknowledges support from the NIHR Comprehensive Biomedical Research Centre, Imperial College Healthcare NHS Trust. The Rotterdam Study (RS) was supported from grants from the Netherlands Organisation for Scientific Research (NWO) Investments (175.010.2005.011, 911-03-012); the Research Institute for Diseases in the Elderly (014-93-015; RIDE2); the Netherlands Genomics Initiative (NGI)/NWO (050-060-810); Erasmus Medical Center; Erasmus University, Rotterdam; Netherlands Organization for the Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture, and Science; the Ministry for Health, Welfare, and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. SAPALDIA was supported by Swiss National Science Foundation grants (no. 3347CO-108796, 3247BO-104283, 3247BO-104288, 3247BO-104284, 32-65896.01, 32-59302.99, 32-52720.97, 32-4253.94, 4026-28099, PDFMP3-123171). Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) is also supported by the Federal Office for Forest, Environment, and Landscape; the Federal Office of Public Health; the Federal Office of Roads and Transport; the canton's government of Aargau, Basel-Stadt, Basel-Land, Geneva, Luzern, Ticino, and Zurich; the Swiss Lung League; and the canton's Lung League of Basel Stadt/Basel Landschaft, Geneva, Ticino, and Zurich. Study of Health in Pomerania (SHIP) is part of the Community Medicine Research net of the University of Greifswald, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the German Asthma and COPD Network (COSYCONET; BMBP grant 01GI0883), the Ministry of Cultural Affairs, as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare (Erlangen, Germany) and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the "Center of Knowledge Interchange" program of the Siemens AG. The TwinsUK study authors acknowledge funding from the Wellcome Trust, the European Community's FP7 (HEALTH-F2-2008-201865-GEFOS), European Network of Genetic and Genomic Epidemiology (ENGAGE) (HEALTH-F4-2007-201413), the FP-5 GenomeEUtwin Project (QLG2-CT-2002-01254), and NIH/National Eye Institute grant 1R01EY018246. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: london2@niehs.nih.gov (S London); mt47@leicester.ac.uk (M Tobin)

† These authors contributed equally to this work.

‡ These authors were joint senior authors on this work.

## Introduction

Spirometric measures of pulmonary function, particularly forced expiratory volume in one second (FEV<sub>1</sub>) and its ratio to forced vital capacity (FEV<sub>1</sub>/FVC), are important clinical tools for diagnosing pulmonary disease, classifying its severity, and evaluating its progression over time. These measures also predict other morbidities and mortality in the general population [1–3]. Genetic factors likely play a prominent role in determining the maximal level of pulmonary function in early adulthood and its subsequent decline with age [4,5]. A relatively uncommon deficiency of  $\alpha$ -1 antitrypsin, due to homozygous mutations of the *SERPINA1* gene, is a well-established genetic risk factor for accelerated decline in pulmonary function, but it accounts for little of the population variability in pulmonary function.

Genome-wide association studies (GWAS) have identified many common genetic variants underlying pulmonary function. The first GWAS of pulmonary function implicated *HHIP* for FEV<sub>1</sub>/FVC [6,7]. GWAS meta-analyses for FEV<sub>1</sub>/FVC and FEV<sub>1</sub> from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and SpiroMeta Consortia have together identified 26 additional novel loci in or near the following genes:

*ADAM19*, *AGER-PPT2*, *ARMC2*, *C10orf11*, *CCDC38*, *CDC123*, *CFDP1*, *FAM13A*, *GPR126*, *HDAC4*, *HTRA4*, *INTS12-GSTCD-NPNT*, *KCNE2*, *LRP1*, *MECOM (EV1)*, *MFAP2*, *MMP15*, *NCR3*, *PDI1*, *PTCH1*, *RARB*, *SPATA9*, *TGFB2*, *THSD4*, *TNSI*, and *ZKSCAN3* [8–10].

Inhaled pollutants, especially cigarette smoking, can have important adverse effects on pulmonary function. Candidate gene studies have not consistently identified interactions with cigarette smoking in relation to pulmonary function. Despite the importance of smoking and other environmental factors in the etiology of many complex human diseases and traits, few GWAS have incorporated gene-by-environment interactions [11–14]. Meta-analyses are generally necessary to provide sufficient sample size to detect moderate effects, and methods for joint testing of single nucleotide polymorphism (SNP) main effects and SNP-by-environment interactions in the meta-analysis setting have only recently been developed [15,16]. This strategy has the potential to identify novel loci that would not emerge from analyses based on the SNP main or interactive effects alone [15–17]. The well-documented and consistent deleterious effect of cigarette smoking on pulmonary function [18] makes it a good candidate for such an approach, since genetic factors may have heterogeneous effects on pulmonary function depending on smoking exposure. We conducted genome-

## Author Summary

Measures of pulmonary function provide important clinical tools for evaluating lung disease and its progression. Genome-wide association studies have identified numerous genetic risk factors for pulmonary function but have not considered interaction with cigarette smoking, which has consistently been shown to adversely impact pulmonary function. In over 50,000 study participants of European descent, we applied a recently developed joint meta-analysis method to simultaneously test associations of gene and gene-by-smoking interactions in relation to two major clinical measures of pulmonary function. Using this joint method to incorporate genetic main effects plus gene-by-smoking interaction, we identified three novel gene regions not previously related to pulmonary function: (1) *DNER*, (2) *HLA-DQB1* and *HLA-DQA2*, and (3) *KCNJ2* and *SOX9*. Expression analyses in human lung tissue from ours or prior studies indicate that these regions contain genes that are plausibly involved in pulmonary function. This work highlights the utility of employing novel methods for incorporating environmental interaction in genome-wide association studies to identify novel genetic regions.

wide joint meta-analyses (JMA) of SNP and SNP-by-smoking interaction (ever-smoking or pack-years) associations with cross-sectional pulmonary function measures ( $FEV_1/FVC$  and  $FEV_1$ ) in 50,047 study participants of European ancestry.

## Results

Table S1 presents characteristics of the 50,047 participants from 19 studies contributing to our analyses. As expected, mean  $FEV_1$  and FVC values were lower in studies with the oldest participants. Standardized residuals of  $FEV_1$  and  $FEV_1/FVC$  (see Methods) were used as the phenotypes for the JMA, in order to maximize comparability with our recent GWAS meta-analysis from the CHARGE and SpiroMeta Consortia [10]. Our original GWAS meta-analyses, conducted separately in CHARGE and SpiroMeta, showed that we were able to identify replicable genetic loci whether using actual pulmonary function measures [8] or their standardized residuals [9]. The standardized residual approach was similarly taken in GWAS of other complex quantitative traits, such as height and body mass index from the Genetic Investigation of ANthropometric Traits (GIANT) Consortium [19,20].

In each of the 19 studies, four regression models with differing SNP-by-smoking interaction terms were run: (1) SNP-by-ever-smoking for standardized  $FEV_1/FVC$  residuals, (2) SNP-by-pack-years for standardized  $FEV_1/FVC$  residuals, (3) SNP-by-ever-smoking for standardized  $FEV_1$  residuals, and (4) SNP-by-pack-years for standardized  $FEV_1$  residuals. Study-specific genomic inflation factors ( $\lambda_{gc}$ ) were calculated for the 1 degree-of-freedom (d.f.) SNP-by-smoking interaction term, to ensure that there was no substantial inflation due to the main effect of smoking being misspecified [21]. All study-specific results had 1 d.f.  $\lambda_{gc} \leq 1.09$  (Table S2), which is of comparable magnitude to other studies with large sample sizes [10,19,22,23].

The study-specific regression coefficients from each of the four models were then combined in JMA, and the resulting  $\lambda_{gc}$  values from the 2 d.f. JMA, calculated across all SNPs, ranged from 1.056 to 1.064. The quantile-quantile plots (Figure S1) show substantial deviation from expectation for SNPs having low  $P$  values from the JMA ( $P_{JMA}$ ). The JMA results corresponding to the top SNP from

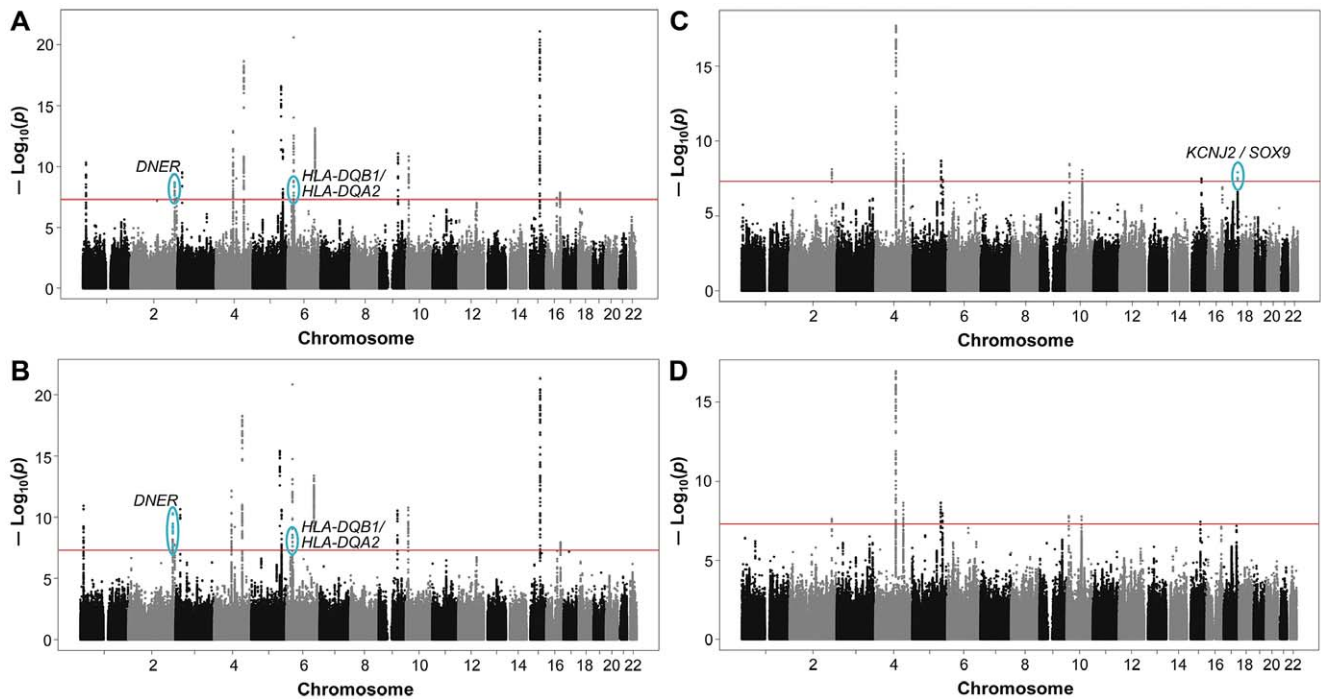
each previously implicated locus [8–10] are presented in Table S3. To identify novel loci among the genome-wide significant loci implicated by our JMA models, the genomic regions surrounding the most significant SNP from each of the 27 previously implicated loci [8–10] (500 kb upstream to 500 kb downstream of each SNP) were removed from consideration (Table S3). Following the removal of all previously implicated loci [8–10], the quantile-quantile plots show that some deviation remained between observed and expected  $P$  values for high-signal SNPs suggesting the presence of novel signals.

In the JMA of SNP and SNP-by-smoking in relation to  $FEV_1/FVC$ , we observed two novel loci containing several significant SNP associations at the standard genome-wide Bonferroni-corrected threshold of  $P_{JMA} < 5 \times 10^{-8}$ , when considering interaction with ever-smoking (Figure 1A) or pack-years (Figure 1B). The SNP associations from both loci also exceeded the more conservative genome-wide significance threshold of  $P_{JMA} < 1.25 \times 10^{-8}$ , based on additional Bonferroni correction for the four JMA models.

The most statistically significant result was for rs7594321, an intronic SNP located in *DNER* (delta/notch-like EGF-related receptor) on chromosome 2, which gave  $P_{JMA} = 2.64 \times 10^{-9}$  (corresponding  $P_{INT} = 0.27$ ) in the ever-smoking model and  $P_{JMA} = 5.00 \times 10^{-11}$  (corresponding  $P_{INT} = 0.0069$ ) in the pack-years model (Table 1). For the ever/never-smoking interaction model, the observed level of significance for the JMA is plausible in the presence of a nominally significant SNP main effect and a nonsignificant interactive effect, as detailed in Text S1. The rs7594321 T allele had a positive  $\beta$  coefficient for the genetic main association and a negative  $\beta$  coefficient for the interaction (Table 1, Table S4 for study-specific results). The regression coefficients correspond to a per allele change of 0.049 (95% CI: 0.030, 0.068) in never-smokers and 0.035 (95% CI: 0.016, 0.053) in ever-smokers. A conserved binding site for the Zic1 transcription factor is located 115 base pairs away from rs7594321. Further, rs7594321 is located upstream of the previously implicated *PIDI* gene (Figure 2A), but it is 713 kb away from the previously implicated SNP (rs1435867), which is located downstream of *PIDI*. There is no linkage disequilibrium (LD) between rs7594321 and rs1435867 ( $r^2 = 0$ ,  $D' = 0$ ).

Our next most statistically significant SNP (rs7764819) is intergenic between two human leukocyte antigen (HLA) genes, *HLA-DQB1* and *HLA-DQA2*, on chromosome 6 (Figure 2B). The *HLA-DQ* region is highly variable, and the association signal in this region is largely driven by two SNPs that are in high LD with one another (rs7764819 and rs7765379,  $r^2 = 1$ ) but only low to moderate LD with all other genotyped and imputed SNPs. A GWAS meta-analysis of asthma implicating the *HLA-DQ* region similarly found highly significant associations with only a few SNPs [24]. Our top SNP rs7764819 gave  $P_{JMA} = 4.39 \times 10^{-9}$  in the ever-smoking model and  $P_{JMA} = 4.35 \times 10^{-9}$  in the pack-years model for  $FEV_1/FVC$  (Table 1). The corresponding  $P_{INT}$  values were  $> 0.05$  (see Text S1). The rs7764819 T allele had negative  $\beta$  coefficients for both the main association and interaction (Table 1, Table S5 for study-specific results), which correspond to a SNP effect of  $-0.060$  (95% CI:  $-0.09$ ,  $-0.031$ ) in never-smokers and  $-0.070$  (95% CI:  $-0.10$ ,  $-0.042$ ) in ever-smokers. Although rs7764819 is located 529 kb away from a previously implicated *AGER* SNP (rs2070600), there is some LD between the two SNPs ( $r^2 = 0.29$ ,  $D' = 0.81$ ). Conserved binding sites for two transcription factors, HTF and Lmo2, are located within 100 kb of rs7764819.

Besides the *DNER* and *HLA-DQB1/HLA-DQA2* loci, SNPs from 12 other chromosomal regions having  $P_{JMA}$  values between



**Figure 1. Genome-wide joint meta-analysis (JMA) of SNP and SNP-by-smoking interaction in relation to pulmonary function.** The Manhattan plots show the chromosomal position of SNPs in comparison to their  $-\log_{10} P_{\text{JMA}}$  values. JMA results are shown for models with (A) SNP-by-ever-smoking interaction term in relation to  $\text{FEV}_1/\text{FVC}$ , (B) SNP-by-pack-years interaction term in relation to  $\text{FEV}_1/\text{FVC}$ , (C) SNP-by-ever-smoking interaction term in relation to  $\text{FEV}_1$ , and (D) SNP-by-pack-years interaction term in relation to  $\text{FEV}_1$ . SNPs located within previously implicated loci are shown, but these loci were not considered when identifying novel loci from the joint modeling of SNP main effects and smoking interactive effects. Novel loci on chromosomes 2, 6, and 17 (shown in blue and circled) were identified as those having SNPs with genome-wide significant  $P$  values at the standard threshold ( $P < 5 \times 10^{-8}$ ) as indicated by the solid red line. Names of the novel gene (or closest genes) are provided. doi:10.1371/journal.pgen.1003098.g001

$5 \times 10^{-8}$  and  $1 \times 10^{-6}$  from either smoking model in relation to  $\text{FEV}_1/\text{FVC}$  are presented in Table S6. Secondary meta-analyses of the interaction product terms alone identified no SNP-by-smoking (ever-smoking or pack-years) interactions at genome-wide statistical significance with  $\text{FEV}_1/\text{FVC}$ . SNPs from two chromosomal regions had  $P_{\text{INT}}$  values between  $5 \times 10^{-8}$  and  $1 \times 10^{-6}$  in relation to  $\text{FEV}_1/\text{FVC}$ , as shown in Table S7.

For  $\text{FEV}_1$ , the JMA of SNP and SNP-by-smoking gave genome-wide significant associations ( $P_{\text{JMA}} < 5 \times 10^{-8}$ ) in the ever-smoking model for four SNPs on chromosome 17 (Figure 1C). However, these SNP associations did not exceed the more conservative significance threshold of  $P_{\text{JMA}} < 1.25 \times 10^{-8}$ . No novel loci reached genome-wide significance level in the pack-years model in relation to  $\text{FEV}_1$  (Figure 1D).

The most significant SNP (rs11654749) from both smoking models is intergenic between *KCNJ2* (a potassium inwardly-rectifying channel also known as KIR2.1) and *SOX9* (sex determining region Y-box 9) (Figure 2C). Conserved binding sites for four transcription factors (HNF-1, CP2, Cdc5, and FOXF2) are located within 100 kb upstream or downstream of rs11654749. The rs11654749 SNP gave  $P_{\text{JMA}} = 1.28 \times 10^{-8}$  in the ever-smoking model and  $P_{\text{JMA}} = 6.63 \times 10^{-8}$  in the pack-years model (Table 1). The corresponding  $P_{\text{INT}}$  values were  $> 0.05$  (see Text S1). The rs11654749 T allele had negative  $\beta$  coefficients for both the main association and interaction (Table 1, Table S8 for study-specific results). These estimates correspond to a SNP effect of  $-0.028$  (95% CI:  $-0.047, -0.010$ ) in never-smokers and  $-0.046$  (95% CI:  $-0.063, -0.029$ ) in ever-smokers. To better understand the magnitude of these  $\beta$  estimates, we compared our results with

those observed in one of our previous GWAS meta-analyses of SNP main effects [9], where standardized residuals of the pulmonary function measures were similarly computed. For a SNP with MAF around 40%, an absolute  $\beta$  value of 0.028 would be equivalent to 19 mL per copy of the risk allele (comparable to a year of  $\text{FEV}_1$  decline in healthy never-smokers), and an absolute  $\beta$  value of 0.046 would be equivalent to 31 mL per copy of the risk allele (comparable to a year and a half of  $\text{FEV}_1$  decline in healthy never-smokers) [25].

Besides this *KCNJ2/SOX9* locus, SNPs from five other chromosomal regions have  $P_{\text{JMA}}$  values between  $5 \times 10^{-8}$  and  $1 \times 10^{-6}$  from either smoking model in relation to  $\text{FEV}_1$  as shown in Table S6. In secondary meta-analyses of the interaction product terms, there were no SNP-by-smoking (ever-smoking or pack-years) interactions implicated at genome-wide statistical significance with  $\text{FEV}_1$ . SNPs from four chromosomal regions had  $P_{\text{INT}}$  values between  $5 \times 10^{-8}$  and  $1 \times 10^{-6}$  in relation to  $\text{FEV}_1$ , as shown in Table S7.

None of the most significant SNPs from the three novel loci we identified by the JMA were associated with  $\text{FEV}_1/\text{FVC}$  or  $\text{FEV}_1$  at or near genome-wide significance in our previous GWAS meta-analysis of 48,201 participants from the CHARGE and SpiroMeta Consortia. In fact, the lowest  $P$  value observed for these SNPs was  $1.04 \times 10^{-3}$  (Table 2) [10].

To evaluate whether the three novel loci identified by the JMA were related to smoking, we evaluated their SNP associations with ever-smoking and cigarettes per day using GWAS meta-analysis results from the Oxford-GlaxoSmithKline (Ox-GSK) Consortium ( $N = 41,150$ ) [26]. None of our implicated SNPs were associated

**Table 1.** Genome-wide significant SNPs from the joint meta-analysis (JMA) of SNP and SNP-by-smoking (ever-smoking or pack-years) interaction in relation to pulmonary function.

SNP (coded allele)	Chr	Gene/closest gene(s)	Coded allele frequency <sup>1</sup>	JMA results							
				Smoking metric	$\beta_{\text{SNP}}^2$	SE <sub>SNP</sub>	$P_{\text{SNP}}$	$\beta_{\text{INT}}^3$	SE <sub>INT</sub>	$P_{\text{INT}}$	$P_{\text{JMA}}$
<i>SNPs implicated in relation to FEV<sub>1</sub>/FVC</i>											
rs7594321 (T)	2q36.3	<i>DNER</i>	0.35	Ever-smoking	0.049	0.0097	$4.14 \times 10^{-7}$	-0.015	0.013	0.27	$2.64 \times 10^{-9}$
				Pack-years	0.048	0.0070	$7.03 \times 10^{-12}$	-0.00020	0.000074	$6.88 \times 10^{-3}$	$5.00 \times 10^{-11}$
rs7764819 (T)	6p21.32	<i>HLA-DQB1/HLA-DQA2</i>	0.89	Ever-smoking	-0.060	0.015	$6.32 \times 10^{-5}$	-0.0010	0.021	0.63	$4.39 \times 10^{-9}$
				Pack-years	-0.064	0.011	$5.95 \times 10^{-9}$	-0.000058	0.00010	0.56	$4.35 \times 10^{-9}$
<i>SNPs implicated in relation to FEV<sub>1</sub></i>											
rs11654749 (T)	17q24.3	<i>KCNJ2/SOX9</i>	0.39	Ever-smoking	-0.028	0.0094	$2.46 \times 10^{-3}$	-0.017	0.013	0.17	$1.28 \times 10^{-8}$
				Pack-years	-0.038	0.0068	$2.29 \times 10^{-8}$	0.000047	0.000068	0.49	$6.63 \times 10^{-8}$

After removing SNPs with known associations with FEV<sub>1</sub>/FVC or FEV<sub>1</sub>, three novel loci with genome-wide significant SNPs (standard threshold of  $P < 5 \times 10^{-8}$ ) remained from the JMA testing in the current study. The most significant SNP from each locus is shown.

FEV<sub>1</sub>, forced expiratory volume in the first second; FVC, forced vital capacity; JMA, joint meta-analysis; SE, standard error; SNP, single nucleotide polymorphism.

<sup>1</sup>Weighted average coded allele frequency across the 19 studies. The coded allele refers to the effect allele.

<sup>2</sup> $\beta_{\text{SNP}}$ , per allele change in the FEV<sub>1</sub>/FVC standardized residual due to the SNP main association.

<sup>3</sup> $\beta_{\text{INT}}$ , per allele change in the FEV<sub>1</sub>/FVC standardized residual due to the interaction between SNP and smoking.

doi:10.1371/journal.pgen.1003098.t001

with these smoking phenotypes at  $P < 0.05$  (Table S9), adding confidence that our JMA-implicated SNP associations were not simply reflective of smoking main effects.

### Expression analyses

Three genes (*DNER*, *KCNJ2*, and *SOX9*) harboring or flanking novel genome-wide significant SNPs were selected for follow-up mRNA expression profiling in human lung tissue and a series of primary cells. Transcripts of all three genes were found in lung tissue, airway smooth muscle, and bronchial epithelial cells; *DNER* and *KCNJ2* transcripts were also found in peripheral blood cells (Table S10).

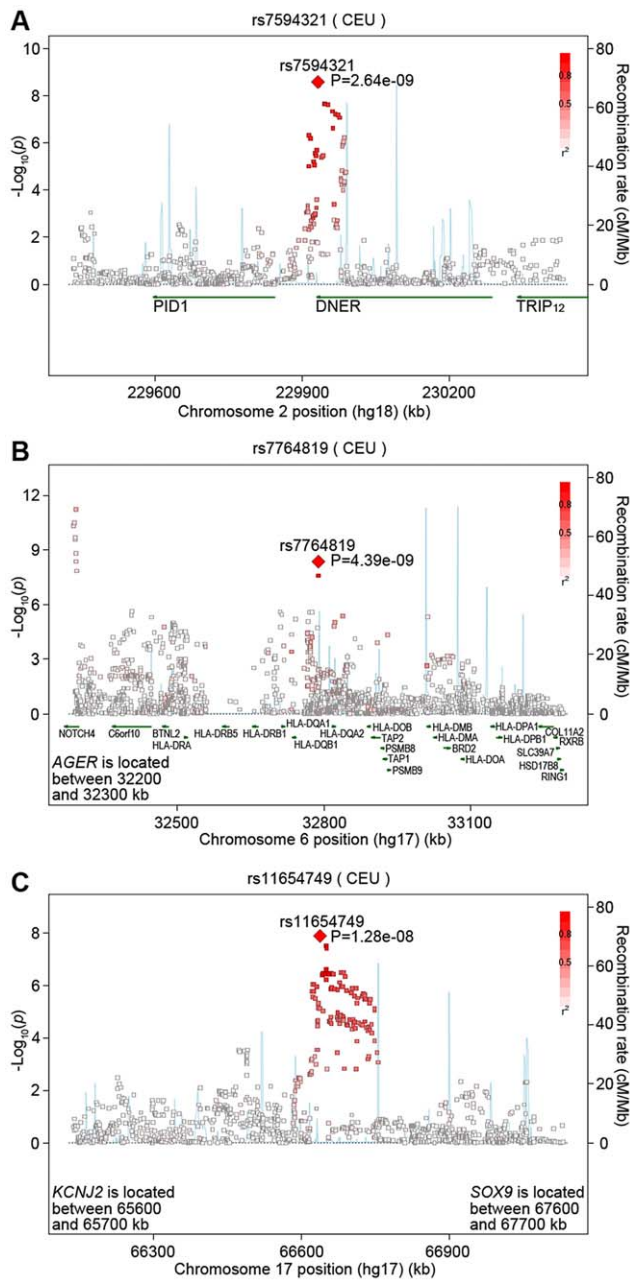
In a separate line of investigation, using the publically available Gene Expression Omnibus repository [27,28], we found that the expression profiling of *DNER* and *SOX9* showed differential expression in human airway epithelium of smokers compared to non-smokers (Figure S2A and S2B) [29]. Expression profiling of *KCNJ2* did not show statistically significant differential expression by smoking status (Figure S2C) [29]. We also identified novel genome-wide significant SNPs in the *HLA-DQ* region, but we did not examine *HLA-DQ* expression given the known expression of class II MHC antigens on a range of airway cell types [30,31]. However, the lead SNP in this region (rs7764819) was associated with statistically significant effects on *HLA-DQB1* expression ( $P = 1.2 \times 10^{-14}$ ), according to an eQTL analysis database of lymphoblastoid cell lines [32].

### Discussion

Few GWAS have accounted for potential interaction with environmental risk factors. To identify novel genetic risk factors that are missed when considering only genetic main effects [33], we used the newly available JMA method [15] to simultaneously summarize regression coefficients for the main SNP and SNP-by-smoking interactive effects in 50,047 participants from 19 studies, based on models that were fully saturated for the main effect of

smoking. This study represents the most comprehensive analysis to date of gene-by-smoking interaction in relation to pulmonary function. We identified two novel loci (*DNER* and *HLA-DQB1/HLA-DQA2*) having highly significant evidence for association with FEV<sub>1</sub>/FVC. A third novel locus (*KCNJ2/SOX9*) was associated with FEV<sub>1</sub>. For the most significant SNPs at each of these three loci, there was no evidence for heterogeneity across the studies (smallest heterogeneity  $P = 0.59$ ), indicating that the associations were not driven by one or a few studies and thus reflect accumulation of evidence across the studies. None of these three loci had previously been associated with pulmonary function. The comparison of results with our prior GWAS meta-analysis of SNP main effects [10], using a comparable sample size, suggested that the SNP associations for our top SNPs were weaker in our previous analyses that examined only genetic main effects. However, our analyses and those of Manning et al. [14] suggest that some of the benefit of using the joint test for some findings comes from the careful adjustment for the environmental main effect. Thus, future studies aimed at replicating these findings may wish to jointly test the SNP main and interactive effects [15,16,33] instead of implementing a standard test of only the SNP main effects. If there is no evidence for interaction at a given locus, the saturation of the main effect of the environmental factor may be important. The joint testing is applicable for both candidate gene [15] and genome-wide [14] approaches. Further, there was minimal overlap in the top SNPs associated with FEV<sub>1</sub>/FVC and FEV<sub>1</sub>, as similarly observed in our previous GWAS meta-analyses of SNP main effects [8–10]. Given that the biological underpinnings of these discrepant association findings remain unknown, future studies should evaluate these genetic loci in the context of the pulmonary function measure for which they were originally implicated.

Given that pulmonary function is a phenotype for which numerous genetic loci have been identified in GWAS and smoking is clearly associated with pulmonary function, it might seem surprising that none of the genome-wide significant SNPs



**Figure 2. Regional association plots of novel loci implicated for pulmonary function.** Three novel loci contained SNPs associated with FEV<sub>1</sub>/FVC or FEV<sub>1</sub> at the standard genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ) in joint meta-analyses of SNP and SNP-by-smoking interaction. SNPs are shown within 500 kb of the most significant SNPs on chromosomes (A) 2q36.3 associated with FEV<sub>1</sub>/FVC, (B) 6p21.32 associated with FEV<sub>1</sub>/FVC, and (C) 17q24.3 associated with FEV<sub>1</sub>. Pairwise  $r^2$  values were based on the HapMap CEU population, and progressively darker shades of red indicate higher  $r^2$  values. Estimated recombination rates from HapMap are shown as background lines.

doi:10.1371/journal.pgen.1003098.g002

implicated by the JMA demonstrated a substantial interaction *per se*. The lack of strong interactive effects does not negate the well-established harmful effects of cigarette smoking nor the need for broad public health campaigns to curb smoking. Instead, our findings demonstrate the value of applying the newly developed

joint methods to uncover novel genetic risk factors that might shed light on the mechanisms leading to reduced pulmonary function.

Our pattern of SNP main and interactive results resemble the patterns seen in another recent application of the same JMA method to incorporate the interaction with body mass index (BMI) into GWAS of type 2 diabetes traits (fasting insulin and blood glucose) [14]. In that study with a sample size of 96,453, nearly double that of ours, the top JMA finding had a corresponding interaction  $P$  value of  $1.6 \times 10^{-4}$  [14]. In our study, the smallest interaction  $P$  value for our top JMA finding was  $6.9 \times 10^{-3}$ . In both our GWAS of smoking and pulmonary function and the recent GWAS of BMI and diabetes traits [14], the SNPs newly implicated by the JMA had marginally significant associations with the trait under study in models with no interaction term, but they became genome-wide significant when accounting for the environmental factor (cigarette smoking or BMI) and the SNP-by-environment interaction. Our JMA included careful modeling of the environmental factor to saturate the environmental main effects along with the interaction testing. In the GWAS of diabetes traits [14], the careful modeling of the environmental factor appeared to account for some of the novel findings from the JMA, consistent with the modest evidence for interaction [14]. Although our previous GWAS meta-analysis was conducted in ever/never-smoking strata, the regression models were not adjusted for smoking status or pack-years [10]. Some of our novel JMA findings compared with our previous GWAS findings may reflect, in part, the saturated modeling of the smoking main effect rather than the interaction *per se*.

The current analysis of 50,047 participants included only 1,846 more participants than our previous GWAS meta-analysis of SNP main effects [10]. To evaluate the likelihood that this 3.8% increase in sample size above that in our previous meta-analysis of pulmonary function was sufficient to explain our identification of these three novel loci at genome-wide statistical significance in the current JMA, we calculated the statistical power to detect genetic main associations (QUANTO [34]) with minor allele frequency (MAF) and  $\beta$  estimates comparable to the three genome-wide significant SNPs presented in Table 1. The current study (total  $N = 50,047$  participants) had only 0.7% to 4.2% more statistical power than our previous GWAS meta-analysis (total  $N = 48,201$  participants) [10], suggesting that the JMA-implicated SNPs are not merely reflective of increased power to detect genetic main effects. Instead, our novel JMA findings demonstrate an advantage of the method used to jointly test the SNP and SNP-by-smoking interactive effects, including the benefit of the saturated modeling of the smoking main effect.

SNPs located in the *DNER* gene were significantly associated with FEV<sub>1</sub>/FVC, even at the more conservative  $P$  value threshold of  $1.25 \times 10^{-8}$ . The JMA results for *DNER* SNPs were driven by both smoking-adjusted main effects and interaction with quantitative smoking history. The *DNER* protein product is a ligand of the Notch signaling pathway that has been implicated in neuronal differentiation and maturation [35,36], adipogenesis [37], and hair-cell development [38]. The Notch pathway is a critical controller of cellular differentiation in multiple organs including the lung [39,40]. Interestingly, the expression levels of many members of the Notch signaling cascade are significantly altered in airway epithelial cells of smokers [41]. We confirmed the expression of *DNER* transcripts in lung and peripheral cells, and by mining publicly available transcriptional profiling databases [29], we found that *DNER* is expressed in bronchial epithelial cells of non-smoking adults and, importantly, its expression is significantly higher in smokers (Figure S2A). Collectively, these results suggest that *DNER* plays a role in cigarette smoke-induced

**Table 2.** Look-up evaluation of SNP main associations with FEV<sub>1</sub>/FVC and FEV<sub>1</sub> using data generated by our previous genome-wide association study meta-analysis (N = 48,201), for the most significant SNP from each of the three novel loci implicated at genome-wide significance in the joint meta-analysis.

SNP (coded allele)	Gene/closest gene(s)	FEV <sub>1</sub> /FVC			FEV <sub>1</sub>		
		$\beta^1$	SE	P	$\beta^1$	SE	P
rs7594321 (T)	<i>DNER</i>	0.032	0.0072	1.04 × 10 <sup>-5</sup>	0.0081	0.0074	0.27
rs7764819 (T)	<i>HLA-DQB1/HLA-DQA2</i>	-0.044	0.011	8.79 × 10 <sup>-5</sup>	-0.0073	0.011	0.52
rs11654749 (T)	<i>KCNJ2/SOX9</i>	-0.023	0.0071	0.0015	-0.031	0.0072	1.23 × 10 <sup>-5</sup>

FEV<sub>1</sub>, forced expiratory volume in the first second; FVC, forced vital capacity; SE, standard error; SNP, single nucleotide polymorphism.

<sup>1</sup> $\beta_{SNP}$ , per allele change in the FEV<sub>1</sub>/FVC standardized residual due to the SNP main association.

doi:10.1371/journal.pgen.1003098.t002

airflow obstruction and further corroborate the importance of the Notch signaling circuitry in the pathogenesis of obstructive lung disease.

Also in relation to FEV<sub>1</sub>/FVC, intergenic SNPs between *HLA-DQB1* and *HLA-DQA2* exceeded the more conservative genome-wide significance threshold. The eQTL analyses indicated that the lead SNP is associated with expression of *HLA-DQB1* specifically. However, the major histocompatibility complex region is highly polymorphic with complex LD patterns, and a few specific functional SNPs might explain the observed associations [42]. Genetic variations within this region have been associated with several autoimmune disorders [43] and asthma [24,44,45], and an interaction between HLA variants and cigarette smoking has been previously implicated [46]. We found little evidence for interaction with smoking at this locus, suggesting that the JMA results were primarily driven by smoking-adjusted genetic main effects. It is most likely that this locus was not identified in our previous GWAS meta-analysis, because the genetic main associations were not evaluated with careful adjustment for smoking status and pack-years. Adjustment for smoking in the current analysis may have removed residual variance in the outcome that is not attributable to genetic variation [14], thus making the identification of the newly associated SNPs possible.

Intergenic SNPs between *KCNJ2* and *SOX9* were significantly associated with FEV<sub>1</sub> at the standard *P* value threshold, but not the more conservative threshold. Similar to the HLA region, it appears that the JMA results for the *KCNJ2/SOX9* region were primarily driven by smoking-adjusted genetic main effects. This region is enriched for long-range regulatory elements for *SOX9*, although the possibility of this region containing *KCNJ2* regulatory elements cannot be discounted [47]. *KCNJ2* is a member of the inwardly-rectifying potassium channel family, which regulates membrane potential and cell excitability and is expressed in many tissues including myocardium, neurons, and vasculature. This potassium channel also affects human bronchial smooth muscle tone and airflow limitation [48]. Dominant negative mutations in *KCNJ2* cause the Andersen syndrome, characterized by ventricular arrhythmias, periodic paralysis, and a number of skeletal and cardiac abnormalities [49]. *SOX9* is a transcription factor that is essential for cartilage formation, [50] but it is also abundantly expressed in other tissues including the respiratory epithelium during development [51]. *Sox9*<sup>-/-</sup> and *Sox9*<sup>+/-</sup> mice have multiple skeletal anomalies and severe tracheal cartilage malformations and die prematurely from respiratory insufficiency [50,52]. Mutations in *SOX9* cause campomelic dysplasia characterized by skeletal defects and autosomal sex reversal [53]. These individuals develop respiratory distress due to chest wall abnormalities, narrowed airways resulting from tracheobronchial defects

and hypoplastic lungs [54]. We confirmed that *KCNJ2* and *SOX9* transcripts were present in human lung tissue and peripheral cells. Using publicly available microarray data [29], we established that *SOX9* is expressed in human airway epithelial cells and its expression is significantly down-regulated in smokers relative to non-smoking adults (Figure S2B). Taken together, these results suggest that *SOX9* may be involved in cigarette smoke-induced airflow obstruction, but further investigation is required to elucidate putative mechanisms.

Most of the previously implicated SNPs had genome-wide significant (or nearly significant) associations with pulmonary function in the JMA, but some were associated with pulmonary function at *P* values that did not approach the genome-wide statistical significance threshold in the JMA analysis. This pattern has two possible explanations. First, the identification of these SNPs at genome-wide statistical significance in our most recent analysis [10] required a sample size of nearly 95,000 individuals, which was obtained by combining discovery and replication cohorts, including additional genotyping on thousands of participants from studies without GWAS data. In the current analysis, the sample size is greatly reduced because of the need for detailed quantitative smoking data and because we were unable to perform additional genotyping in studies without GWAS data. Second, Manning et al. [15] showed that a meta-analysis of main SNP effects has slightly greater power than the JMA under the scenario of no interaction, so it is not surprising that a few of the prior SNP findings had varying levels of significance between our prior GWAS meta-analyses [8–10] and the current JMA study. While our sample size of over 50,000 study participants is large, and the study of Manning et al. [14] examining SNP-by-BMI interaction in relation to fasting insulin is nearly twice as large, identification of interactions is challenging from a statistical power perspective. Given the multiple testing issues in genome interaction testing, even larger sample sizes will likely be needed to identify gene-by-environment interactions with rare variants or with the modest effect sizes that we generally expect. Nonetheless, our findings exemplify the greater power achieved by using the joint methods, such as those reported by Manning et al. [15] and Kraft et al. [16,33], to incorporate interaction with a clearly associated environmental risk factor. The novel genetic loci identified here for pulmonary function would have remained unknown using standard GWAS approaches.

## Methods

### Ethics statement

Nineteen independent studies contributed to our analyses. All study protocols were approved by the respective local Institutional



Review Boards, and written informed consent for genetic studies was obtained from all participants included in our analyses.

### Cohort studies

Of the 19 studies contributing to our analyses, 18 studies came from the CHARGE [8,55] or SpiroMeta [9] Consortium: Age, Gene, Environment, Susceptibility (AGES) – Reykjavik Study [56]; Atherosclerosis Risk in Communities (ARIC) Study [57]; British 1958 Birth Cohort (B58C) [58]; Coronary Artery Risk Development in Young Adults (CARDIA) [59,60]; Cardiovascular Health Study (CHS) [61]; European Community Respiratory Health Survey (ECRHS) [62]; European Prospective Investigation into Cancer and Nutrition (EPIC, obese cases and population-based subsets) [63]; Framingham Heart Study (FHS) [64,65]; Health, Aging, and Body Composition (Health ABC) Study [66]; Northern Finland Birth Cohort of 1966 (NFBC1966) [67,68]; Multi-Ethnic Study of Atherosclerosis (MESA) [69,70]; Rotterdam Study (RS-I, RS-II, and RS-III) [71]; Swiss Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) [72]; Study of Health in Pomerania (SHIP) [73]; and TwinsUK [74]. We reached out to other population-based studies with GWAS genotyping and data available on cigarette smoking and pulmonary function, resulting in the inclusion of LifeLines [75]. Given the greater power needed to detect novel genetic loci with subtle gene-environment interaction regardless of the statistical method used [16], we chose to maximize statistical power to discover novel genetic loci by combining all available participants and to use the regression coefficients across the many different component studies as evidence for consistency. This approach was similarly taken by another large-scale GWAS consortium for discovering SNP main effects [24].

### Pulmonary function measurements and smoking information

All studies were included in our previous GWAS meta-analysis of pulmonary function or the follow-up replication analyses, wherein their pulmonary function testing protocols were described [10]. For studies with spirometry at a single visit (B58C, LifeLines, MESA, NFBC1966, SHIP, RS-I, RS-II, and RS-III), we analyzed FEV<sub>1</sub>/FVC and FEV<sub>1</sub> measured at that visit. For studies with spirometry at more than one visit, we analyzed measurements from the baseline visit (AGES, ARIC, CARDIA, CHS, ECRHS, EPIC obese cases, EPIC population-based, Health ABC, and SAPALDIA) or the most recent examination with spirometry data (FHS and TwinsUK).

Smoking history (current-, past-, and never-smoking) was ascertained by questionnaire at the time of pulmonary function testing. Pack-years of smoking were calculated for current and past smokers by multiplying smoking amount (packs/day) and duration (years smoked). Table S11 presents the specific questions used to ascertain smoking history and pack-years in each of the 19 studies.

### Genotyping, quality control, and imputation

Study participants were genotyped on various genotyping platforms, and standard quality control filters for call rate, Hardy-Weinberg equilibrium p-value, MAF, and other measures were applied to the genotyped SNPs (Table S12). To generate a common set of SNPs for meta-analysis, imputation was conducted with reference haplotype panels from HapMap phase II subjects of European ancestry (CEU) (Table S12) [76]. Imputed genotype dosage values (estimated reference allele count with a fractional value ranging from 0 to 2.0) were generated for approximately 2.5 million autosomal SNPs. Among participants with genome-wide SNP genotyping data, exclusions were made due to standard quality

control metrics (call rate, discordance with prior genotyping, and genotypic and phenotypic sex mismatch among others), missing pulmonary function data, or missing covariate data (Table S13).

### Statistical analysis

Our analyses included 50,047 participants from 19 studies who passed their study-specific quality control and had complete data on pulmonary function and smoking. Each study transformed the pulmonary function measures to residuals using linear regression of FEV<sub>1</sub>/FVC (%) and FEV<sub>1</sub> (mL) on age, age<sup>2</sup>, sex, and standing height as predictors. Principal component eigenvectors and recruitment site were also included as covariates to adjust for population stratification (if applicable). The residuals were converted to z scores (henceforth referred to as standardized residuals). We confirmed that smoking was inversely associated with the FEV<sub>1</sub>/FVC and FEV<sub>1</sub> standardized residuals in all 19 studies (meta-analysis  $\beta = -0.0030$  and corresponding  $P < 1 \times 10^{-6}$  for pack-years of smoking).

The FEV<sub>1</sub>/FVC and FEV<sub>1</sub> standardized residuals were used as the phenotypes for genome-wide association testing with linear regression models, which included the following predictor variables: imputed SNP genotype dosages, smoking history (dichotomous variable, 0 = never-smokers and 1 = ever-smokers), smoking status (dichotomous variable, 0 = never- and past-smokers and 1 = current-smokers), pack-years of smoking (continuous variable), and a SNP-by-smoking interaction product term. Two of the 19 studies (FHS and TwinsUK) had much relatedness among participants, and we took appropriate account of relatedness in the association testing (Table S12). Four regression models with interaction terms for ever-smoking or pack-years were specified in relation to standardized residuals for FEV<sub>1</sub>/FVC or FEV<sub>1</sub>. As it has long been advised in studying interactions, the regression models were designed to fully saturate the main smoking effect on pulmonary function, so that the interaction terms do not capture residual main effects [77]. In each of the 19 studies, the genome-wide analyses were implemented with robust variance estimation using the software packages indicated in Table S12.

Our analyses were aimed at finding novel loci associated with pulmonary function when considering an interaction with cigarette smoking, so we chose to implement JMA of SNP main and interactive SNP-by-smoking effects (two d.f. test of the null hypothesis  $\beta_{\text{SNP}} = 0$  and  $\beta_{\text{INT}} = 0$ ) [15]. Manning et al. previously compared the joint methods, such as JMA, with other methods that incorporate gene-environment interaction (such as screening by main effects [78] or conducting a 1 d.f. meta-analysis of the interaction product term), and they found that the joint methods offer optimal statistical power over a range of scenarios for SNP main and interactive effects [15,33]. Therefore, our analyses centered on the JMA method, which simultaneously estimates regression coefficients for the SNP and SNP-by-smoking interaction terms, while accounting for their covariance, to generate a joint test of significance [15]. It also accounts for the unequal variances from studies of different sample sizes. Secondarily, we implemented meta-analyses of just the  $\beta$  coefficient from the interaction term for comparison with the JMA results. Of note, the two-step gene-environment interaction study designs by Murcay et al. [79,80] and Gauderman et al. [81] are applicable to case-control or case-parent trio studies, respectively, and were thus not considered for our population-based studies of continuous traits.

The JMA was conducted with fixed effects on approximately 2.5 million SNPs using METAL software (version 2010-02-08) [82] and patch source code provided by Manning et al. [82]. Genomic control correction was applied by computing  $\lambda_{\text{gc}}$  as the ratio of the observed and expected (2 d.f.) median chi-square statistics and

dividing the observed chi-square statistics by  $\lambda_{gc}$ . SNPs having  $P_{JMA} < 5 \times 10^{-8}$  (the standard Bonferroni-adjusted  $P$  value) were considered statistically significant [83]. Further correction for the four different (albeit related) JMA models yielded a conservative  $P_{JMA}$  threshold of  $1.25 \times 10^{-8}$ . In addition to reporting the  $P_{JMA}$  for the most significant SNP from each novel locus, we used the  $\beta$  and standard error (SE) estimates from the JMA results to calculate the  $P$  values corresponding to the SNP main association ( $P_{SNP}$ ) and the SNP-by-ever-smoking interaction ( $P_{INT}$ ) [15].

### Bioinformatics analysis

Gene annotation was performed using the gene prediction tracks “UCSC Genes” and “RefSeq Genes” in the UCSC browser (<http://genome.ucsc.edu>). The “sno/miRNA” track from the UCSC browser was used to search for any microRNA within 100 kb upstream or downstream of each SNP, and the “TFBS Conserved” track was used to search for conserved transcription factor binding sites (TFBSs) at or near the most significant SNPs. The SNAP program [84] was used to infer LD patterns, based on the HapMap phase II CEU population.

### Expression analyses

We used separate types of expression analyses to confirm the biologic plausibility of our findings. First, we carried out mRNA expression profiling to show whether or not the implicated genes are expressed in human tissues relevant to pulmonary function. The mRNA expression profiles of implicated genes were determined using reverse transcription polymerase chain reaction (RT-PCR). RNA was sourced from lung (Ambion/ABI), human bronchial epithelial cells (Clonetics) [85], and peripheral blood mononuclear cells (3H Biomedica). RNA from human airway smooth muscle cells, cultured as previously described from tissue obtained at thoracotomy [86], was extracted using a commercially available kit (Qiagen). Ethical approval for the use of primary cells was obtained from the local ethics committees. cDNA was generated using 1  $\mu$ g of RNA template using random hexamers and a SuperScript kit (Invitrogen) as directed by the manufacturer. PCR assays were designed to cross intron-exon boundaries, where possible and where splice variation was known, in order to detect all variants. The *GAPDH* gene was used as a positive control for the cDNA quality, and water was used as a negative control. Primer sequences for the genes of interest are given in Table S14. All PCR were done using Platinum Taq High Fidelity (Invitrogen) with 100 ng of cDNA template in a 25  $\mu$ L reaction. Cycling conditions were as follows: 94°C for 2 minutes, 35 cycles of 94°C for 45 seconds, 55°C for 30 seconds, and 68°C for 90 seconds. Following PCR, gel bands were directly sequenced to confirm the presence of the gene’s transcript.

Second, we used another publically available data repository to investigate whether any of the implicated genes showed evidence for differential expression depending on smoking history. The gene expression profiles of human airway epithelium from healthy smokers ( $N = 10$ ) and nonsmokers ( $N = 12$ ) were obtained from the Gene Expression Omnibus site (<http://www.ncbi.nlm.nih.gov/geo/>) [27,28], based on robust multichip average processing of probe intensities from Affymetrix HG-U133 Plus 2.0 microarrays (GEO dataset number GSE4498) [29]. Mean expression levels of genes around our genome-wide significant findings from the JMA were compared between smokers versus nonsmokers. The  $P$  value for the difference in means between smokers and nonsmokers was calculated using the nonparametric Mann-Whitney test.

Third, our genome-wide significant SNPs from novel loci were searched against an expression quantitative trait loci (eQTL) data repository based on lymphoblastoid cell lines [32], to investigate whether any of the implicated SNP variants might influence the

expression of the nearby genes.  $P < 5 \times 10^{-8}$  was used to designate statistically significant eQTL associations.

### Supporting Information

**Figure S1** Quantile–quantile plots for the genome-wide joint meta-analysis (JMA) of SNP and SNP-by-smoking interaction in relation to pulmonary function. The plots compare the observed vs. expected  $P$  values for JMA testing of SNPs by (A) ever-smoking in relation to FEV<sub>1</sub>/FVC, (B) pack-years of smoking in relation to FEV<sub>1</sub>/FVC, (C) ever-smoking in relation to FEV<sub>1</sub>, and (D) pack-years of smoking in relation to FEV<sub>1</sub>. The corresponding two degree-of-freedom genomic inflation factors ( $\lambda_{gc}$ ) are shown, as calculated across all SNPs before the exclusion of previously implicated SNPs. The JMA results of all SNPs were plotted (in blue), along with the SNPs remaining after exclusion of the 27 previously implicated loci (in black).

(DOCX)

**Figure S2** mRNA expression profiling in human airway epithelium from healthy smokers versus nonsmokers. Expression profiles of 10 smokers (indicated in blue) and 12 nonsmokers (indicated in red) were obtained for (A) *DNER*, (B) *SOX9*, and (C) *KCNJ2*, using microarray data from the Gene Expression Omnibus site (<http://www.ncbi.nlm.nih.gov/geo/>) (GSE4498). The y-axes reflect the probe intensities of each gene transcript from Affymetrix HG-U133 Plus 2.0 microarrays [29], with the horizontal bold bars indicating the average probe intensities and the smaller bars indicating standard deviation. *SOX9* was represented by two different probes on the microarray; therefore, the intensities were averaged for each sample. The  $P$  value was calculated using the nonparametric Mann-Whitney test.

(DOCX)

**Table S1** Characteristics of study participants (total  $N = 50,047$ ) at the time of pulmonary function testing.

(DOCX)

**Table S2** Genomic inflation factors ( $\lambda_{gc}$ ) for study-specific results (corresponding to the 1 degree of freedom SNP-by-smoking product term) in each of the four regression models.

(DOCX)

**Table S3** Regions surrounding the most significant SNP from each of 27 previously implicated loci (500 kb upstream to 500 kb downstream of each SNP). These loci were excluded when identifying novel loci from the joint meta-analysis (JMA) of SNP and SNP-by-smoking interaction. The smallest  $P$  value from the JMA ( $P_{JMA}$ ) is shown, along with the corresponding JMA model from which the result was obtained.

(DOCX)

**Table S4** Study-specific results for the genome-wide significant SNP rs7594321 (coded allele: T), located in the *DNER* gene.  $\beta$  estimates and  $P$  values are shown for the SNP main association ( $\beta_{SNP}$  and  $P_{SNP}$ ) and interactive association ( $\beta_{INT}$  and  $P_{INT}$ ) by smoking (ever-smoking and pack-years) in relation to FEV<sub>1</sub>/FVC. The  $P$  values corresponding to the joint test of SNP main and interactive associations are also shown.

(DOCX)

**Table S5** Study-specific results for the genome-wide significant SNP rs7764819 (coded allele: T), located between the *HLA-DQB1* and *HLA-DQA2* genes.  $\beta$  estimates and  $P$  values are shown for the SNP main association ( $\beta_{SNP}$  and  $P_{SNP}$ ) and interactive association ( $\beta_{INT}$  and  $P_{INT}$ ) by smoking (ever-smoking and pack-years) in relation to FEV<sub>1</sub>/FVC. The  $P$  values corresponding to the joint test of SNP main and interactive associations are also shown.

(DOCX)

**Table S6** SNPs from each of 16 chromosomal regions with  $P$  values between  $5 \times 10^{-8}$  and  $1 \times 10^{-6}$  for the joint meta-analysis of SNP and SNP-by-smoking (ever-smoking or pack-years) in relation to pulmonary function ( $FEV_1/FVC$  or  $FEV_1$ ). A hyphen (“–”) indicates  $P > 1 \times 10^{-6}$ . For each regression model, the SNP having the smallest  $P_{JMA}$  from each locus is shown. (DOCX)

**Table S7** SNPs with  $P < 1 \times 10^{-6}$  from the 1 degree-of-freedom meta-analysis of regression coefficients corresponding to the SNP-by-smoking (ever-smoking or pack-years) interaction term in relation to  $FEV_1/FVC$ . No SNPs exceeded the standard genome-wide significance threshold ( $P < 5 \times 10^{-8}$ ). A hyphen (“–”) indicates  $P > 1 \times 10^{-6}$ . For each regression model, the SNP having the smallest  $P_{INT}$  from each locus is shown. (DOCX)

**Table S8** Study-specific results for the genome-wide significant SNP rs11654749 (coded allele: T), located between the *KCNJ2* and *SOX9* genes.  $\beta$  estimates and  $P$  values are shown for the SNP main association ( $\beta_{SNP}$  and  $P_{SNP}$ ) and interactive association ( $\beta_{INT}$  and  $P_{INT}$ ) by ever-smoking in relation to  $FEV_1$ . The  $P$  values corresponding to the joint test of SNP main and interactive associations are also shown. (DOCX)

**Table S9** Look-up evaluation of main SNP associations with cigarette smoking phenotypes using data generated by the Oxford-GlaxoSmithKline Consortium ( $N = 41,150$ ), for the most significant SNP from each of the three novel loci implicated at genome-wide significance in the joint meta-analysis. (DOCX)

**Table S10** mRNA expression profiling of three candidate genes in the human lung and periphery. Primer sequences are provided in Table S5. A “+” sign indicates presence of the transcript, and “–” indicates its absence. All products were sequence verified. (DOCX)

**Table S11** Questionnaire data used to ascertain cigarette smoking history (ever-smoking), amount, and duration across the 19 studies. Smoking amount and duration were used together to calculate pack-years. (DOCX)

**Table S12** Details of single nucleotide polymorphism (SNP) genotyping, quality control (QC), imputation, and statistical analysis across the 19 studies. (DOCX)

**Table S13** Study participants of European descent and quality control (QC) across the 19 studies. Participants passing QC filters and having acceptable spirometry data and complete covariate data were included in the meta-analyses. (DOCX)

**Table S14** Primers for mRNA expression profiling. (DOCX)

**Text S1** Detailed explanation of joint meta-analysis significance levels, in relation to main and interactive significance. (DOCX)

## Acknowledgments

The authors thank the other investigators, staff, and participants of all cohort studies for their valuable contributions. The authors from Lifelines are also grateful to the Medical Biobank Northern Netherlands and participating general practitioners and pharmacists. A full list of principal Cardiovascular Health Study (CHS) investigators and institutions can be found at <http://www.chs-nhlbi.org/pi.htm>. A full list of participating Multi-Ethnic Study of Atherosclerosis (MESA) investigators and institutions can be found at <http://www.mesa-nhlbi.org>.

## Author Contributions

Conceived and designed the experiments: DB Hancock, MS Artigas, SA Gharib, A Henry, A Manichaikul, A Ramasamy, DW Loth, M Imboden, B Koch, WL McArdle, AV Smith, J Smolonska, A Sood, W Tang, JB Wilk, G Zhai, JH Zhao, H Aschard, KM Burkart, I Curjuric, M Eijgelsheim, P Elliott, X Gu, TB Harris, C Janson, G Homuth, PG Hysi, JZ Liu, LR Loehr, K Lohman, RJF Loos, AK Manning, KD Marcic, M Obeisat, DS Postma, MC Aldritch, GG Brusselle, T-h Chen, G Eiriksdottir, N Franceschini, J Heinrich, JI Rottner, C Wijmenga, OD Williams, AR Bentley, A Hofman, CC Laurie, T Lumley, AC Morrison, BR Joubert, F Rivadeneira, DJ Couper, SB Kritchevsky, Y Liu, M Wjst, LV Wain, JM Vonk, AG Uitterlinden, T Rochat, SS Rich, BM Pstaj, GT O'Connor, KE North, DB Mirel, B Meibohm, LJ Launer, K-T Khaw, A-L Hartikainen, CJ Hammond, S Glaser, J Marchini, P Kraft, NJ Wareham, H Völzke, BHC Stricker, TD Spector, NM Probst-Hensch, D Jarvis, M-R Jarvelin, SR Heckbert, V Gudnason, HM Boezen, RG Barr, PA Cassano, DP Strachan, M Fornage, IP Hall, J Dupuis, MD Tobin, SJ London. Performed the experiments: A Henry, A Hofman. Analyzed the data: DB Hancock, MS Artigas, SA Gharib, A Henry, A Manichaikul, A Ramasamy, DW Loth, M Imboden, WL McArdle, AV Smith, J Smolonska, A Sood, W Tang, JB Wilk, G Zhai, JH Zhao, A Hofman. Contributed reagents/materials/analysis tools: H Aschard, AK Manning, P Kraft, J Dupuis, A Henry, IP Hall. Wrote the paper: DB Hancock, MS Artigas, SA Gharib, A Henry, A Manichaikul, A Ramasamy, DW Loth, PA Cassano, DP Strachan, M Fornage, IP Hall, J Dupuis, MD Tobin, SJ London.

## References

- Hole DJ, Watt GC, Davey-Smith G, Hart CL, Gillis CR, et al. (1996) Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study. *BMJ* 313: 711–716.
- Schunemann HJ, Dorn J, Grant BJ, Winkelstein W, Jr., Trevisan M (2000) Pulmonary function is a long-term predictor of mortality in the general population: 29-year follow-up of the Buffalo Health Study. *Chest* 118: 656–664.
- Myint PK, Luben RN, Surtees PG, Wainwright NW, Welch AA, et al. (2005) Respiratory function and self-reported functional health: EPIC-Norfolk population study. *Eur Respir J* 26: 494–502.
- Redline S, Tishler PV, Lewitter FI, Tager IB, Munoz A, et al. (1987) Assessment of genetic and nongenetic influences on pulmonary function. A twin study. *Am Rev Respir Dis* 135: 217–222.
- Hubert HB, Fabsitz RR, Feinleib M, Gwinn C (1982) Genetic and environmental influences on pulmonary function in adult twins. *Am Rev Respir Dis* 125: 409–415.
- Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, et al. (2009) A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 5: e1000429. doi:10.1371/journal.pgen.1000429
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, et al. (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 5: e1000421. doi:10.1371/journal.pgen.1000421
- Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, et al. (2010) Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 42: 45–52.
- Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, et al. (2010) Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 42: 36–44.
- Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, et al. (2011) Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 43: 1082–1090.
- Beatty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, et al. (2011) Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genet Epidemiol* 35: 469–478.
- Liu Y, Xu H, Chen S, Chen X, Zhang Z, et al. (2011) Genome-wide interaction-based association analysis identified multiple new susceptibility Loci for common diseases. *PLoS Genet* 7: e1001338. doi:10.1371/journal.pgen.1001338
- Ege MJ, Strachan DP, Cookson WO, Moffatt MF, Gut I, et al. (2011) Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol* 127: 138–144, 144 e131–134.

14. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 44: 659–669.
15. Manning AK, LaValley M, Liu C-T, Rice K, An P, et al. (2011) Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP×Environment regression coefficients. *Genet Epidemiol* 35: 11–18.
16. Aschard H, Hancock DB, London SJ, Kraft P (2010) Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered* 70: 292–300.
17. Hamza TH, Chen H, Hill-Burns EM, Rhodes SL, Montimurro J, et al. (2011) Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS Genet* 7: e1002237. doi:10.1371/journal.pgen.1002237
18. (2010). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta (GA).
19. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467: 832–838.
20. Wen W, Cho YS, Zheng W, Dorajoo R, Kato N, et al. (2012) Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* 44: 307–311.
21. Cornelis MC, Tchetgen Tchetgen EJ, Liang L, Qi L, Chatterjee N, et al. (2012) Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *Am J Epidemiol* 175: 191–202.
22. Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, et al. (2010) Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 42: 1077–1085.
23. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, et al. (2009) Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 5: e1000508. doi:10.1371/journal.pgen.1000508
24. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genome-wide association study of asthma. *N Engl J Med* 363: 1211–1221.
25. Kohansal R, Martinez-Camblor P, Agusti A, Buist AS, Mannino DM, et al. (2009) The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* 180: 3–10.
26. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, et al. (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 42: 436–440.
27. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207–210.
28. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res* 39: D1005–1010.
29. Harvey BG, Heguy A, Leopold PL, Carolan BJ, Ferris B, et al. (2007) Modification of gene expression of the small airway epithelium in response to cigarette smoking. *J Mol Med (Berl)* 85: 39–53.
30. Caulfield JJ, Fernandez MH, Sousa AR, Lane SJ, Lee TH, et al. (1999) Regulation of major histocompatibility complex class II antigens on human alveolar macrophages by granulocyte-macrophage colony-stimulating factor in the presence of glucocorticoids. *Immunology* 98: 104–110.
31. Glanville AR, Tazelaar HD, Theodore J, Imoto E, Rouse RV, et al. (1989) The distribution of MHC class I and II antigens on bronchial epithelium. *Am Rev Respir Dis* 139: 330–334.
32. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. *Nat Genet* 39: 1202–1207.
33. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ (2007) Exploiting gene-environment interaction to detect genetic associations. *Hum Hered* 63: 111–119.
34. Gauderman WJ, Morrison JM (2006) QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.
35. Tohgo A, Eiraku M, Miyazaki T, Miura E, Kawaguchi SY, et al. (2006) Impaired cerebellar functions in mutant mice lacking DNER. *Mol Cell Neurosci* 31: 326–333.
36. Fukazawa N, Yokoyama S, Eiraku M, Kengaku M, Maeda N (2008) Receptor type protein tyrosine phosphatase zeta-pleiotrophin signaling controls endocytic trafficking of DNER that regulates neurogenesis. *Mol Cell Biol* 28: 4494–4506.
37. Park JR, Jung JW, Seo MS, Kang SK, Lee YS, et al. (2010) DNER modulates adipogenesis of human adipose tissue-derived mesenchymal stem cells via regulation of cell proliferation. *Cell Prolif* 43: 19–28.
38. Kowalik L, Hudspeth AJ (2011) A search for factors specifying tonotopy implicates DNER in hair-cell development in the chick's cochlea. *Dev Biol* 354: 221–231.
39. Guseh JS, Bores SA, Stanger BZ, Zhou Q, Anderson WJ, et al. (2009) Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. *Development* 136: 1751–1759.
40. Tsao PN, Vasconcelos M, Izvolyski KI, Qian J, Lu J, et al. (2009) Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. *Development* 136: 2297–2307.
41. Tilley AE, Harvey BG, Heguy A, Hackett NR, Wang R, et al. (2009) Down-regulation of the notch pathway in human airway epithelium in association with smoking and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 179: 457–466.
42. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 44: 291–296.
43. Zhernakova A, van Diemen CC, Wijmenga C (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet* 10: 43–55.
44. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, et al. (2011) Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat Genet* 43: 893–896.
45. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, et al. (2011) Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* 43: 887–892.
46. Mahdi H, Fisher BA, Kallberg H, Plant D, Malmstrom V, et al. (2009) Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nat Genet* 41: 1319–1324.
47. Gordon CT, Tan TY, Benko S, Fitzpatrick D, Lyonnet S, et al. (2009) Long-range regulation at the SOX9 locus in development and disease. *J Med Genet* 46: 649–656.
48. Oonuma H, Iwasawa K, Iida H, Nagata T, Imuta H, et al. (2002) Inward rectifier K(+) current in human bronchial smooth muscle cells: inhibition with antisense oligonucleotides targeted to Kir2.1 mRNA. *Am J Respir Cell Mol Biol* 26: 371–379.
49. Andelfinger G, Tapper AR, Welch RC, Vandoye CG, George AL, Jr., et al. (2002) KCNJ2 mutation results in Andersen syndrome with sex-specific cardiac and skeletal muscle phenotypes. *Am J Hum Genet* 71: 663–668.
50. Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrughe B (1999) Sox9 is required for cartilage formation. *Nat Genet* 22: 85–89.
51. Liu Y, Hogan BL (2002) Differential gene expression in the distal tip endoderm of the embryonic mouse lung. *Gene Expr Patterns* 2: 229–233.
52. Bi W, Huang W, Whitworth DJ, Deng JM, Zhang Z, et al. (2001) Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc Natl Acad Sci U S A* 98: 6698–6703.
53. Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, et al. (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372: 525–530.
54. Houston CS, Opitz JM, Spranger JW, Macpherson RI, Reed MH, et al. (1983) The campomelic syndrome: review, report of 17 cases, and follow-up on the currently 17-year-old boy first reported by Maroteaux et al in 1971. *Am J Med Genet* 15: 3–28.
55. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, et al. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet* 2: 73–80.
56. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, et al. (2007) Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* 165: 1076–1087.
57. (1989) The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 129: 687–702.
58. Strachan DP, Rudnicka AR, Power C, Shepherd P, Fuller E, et al. (2007) Lifecourse influences on health among British adults: effects of region of residence in childhood and adulthood. *Int J Epidemiol* 36: 522–531.
59. Hughes GH, Cutter G, Donahue R, Friedman GD, Hulley S, et al. (1987) Recruitment in the Coronary Artery Disease Risk Development in Young Adults (Cardia) Study. *Control Clin Trials* 8: 68S–73S.
60. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, et al. (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* 41: 1105–1116.
61. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, et al. (1991) The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1: 263–276.
62. Burney PG, Luczynska C, Chinn S, Jarvis D (1994) The European Community Respiratory Health Survey. *Eur Respir J* 7: 954–960.
63. Day N, Oakes S, Luben R, Khaw KT, Bingham S, et al. (1999) EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer. Br J Cancer* 80 Suppl 1: 95–103.
64. Dawber TR, Kannel WB (1966) The Framingham study. An epidemiological approach to coronary heart disease. *Circulation* 34: 553–555.
65. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP (1975) The Framingham Offspring Study. Design and preliminary data. *Prev Med* 4: 518–525.
66. Yende S, Waterer GW, Tolley EA, Newman AB, Bauer DC, et al. (2006) Inflammatory markers are associated with ventilatory limitation and muscle dysfunction in obstructive lung disease in well functioning elderly subjects. *Thorax* 61: 10–16.
67. Jarvelin MR, Hartikainen-Sorri AL, Rantakallio P (1993) Labour induction policy in hospitals of different levels of specialisation. *Br J Obstet Gynaecol* 100: 310–315.
68. Rantakallio P (1969) Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr Scand* 193: Suppl 193:191+.

69. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, et al. (2002) Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 156: 871–881.
70. Rodriguez J, Jiang R, Johnson WC, MacKenzie BA, Smith LJ, et al. (2010) The association of pipe and cigar use with cotinine levels, lung function, and airflow obstruction: a cross-sectional study. *Ann Intern Med* 152: 201–210.
71. Hofman A, van Duijn CM, Franco OH, Ikram MA, Janssen HL, et al. (2011) The Rotterdam Study: 2012 objectives and design update. *Eur J Epidemiol* 26: 657–686.
72. Martin BW, Ackermann-Liebrich U, Leuenberger P, Kunzli N, Stutz EZ, et al. (1997) SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. *Soz Präventivmed* 42: 67–84.
73. Volzke H, Alte D, Schmidt CO, Radke D, Lörbe R, et al. (2011) Cohort profile: the study of health in Pomerania. *Int J Epidemiol* 40: 294–307.
74. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, et al. (2001) Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* 4: 464–477.
75. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, et al. (2008) Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur J Epidemiol* 23: 67–74.
76. (2003) The International HapMap Project. *Nature* 426: 789–796.
77. Shahani AK (1970) A Saturated Experiment in Sequential Determination of Operating Conditions. *Journal of the Royal Statistical Society Series D (The Statistician)* 19: 403–408.
78. Kooperberg C, LeBlanc M (2008) Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol* 32: 255–263.
79. Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ (2011) Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genet Epidemiol* 35: 201–210.
80. Murcray CE, Lewinger JP, Gauderman WJ (2009) Gene-environment interaction in genome-wide association studies. *Am J Epidemiol* 169: 219–226.
81. Gauderman WJ, Thomas DC, Murcray CE, Conti D, Li D, et al. (2010) Efficient genome-wide association testing of gene-environment interaction in case-parent trios. *Am J Epidemiol* 172: 116–122.
82. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 26: 2190–2191.
83. Pe'er I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. *Genet Epidemiol* 32: 381–385.
84. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24: 2938–2939.
85. Wadsworth SJ, Nijmeh HS, Hall IP (2006) Glucocorticoids increase repair potential in a novel in vitro human airway epithelial wounding model. *J Clin Immunol* 26: 376–387.
86. Sayers I, Swan C, Hall IP (2006) The effect of beta2-adrenoceptor agonists on phospholipase C (beta1) signalling in human airway smooth muscle cells. *Eur J Pharmacol* 531: 9–12.