

University of Groningen

## Susceptibility to chronic mucus hypersecretion, a genome wide association study

Dijkstra, Akkelies E.; Smolonska, Joanna; van den Berge, Maarten; Wijmenga, Cisca; Zanen, Pieter; Luinge, Marjan A.; Platteel, Mathieu; Lammers, Jan-Willem; Dahlback, Magnus; Tosh, Kerrie

*Published in:*  
 PLoS ONE

*DOI:*  
[10.1371/journal.pone.0091621](https://doi.org/10.1371/journal.pone.0091621)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
 Publisher's PDF, also known as Version of record

*Publication date:*  
 2014

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Dijkstra, A. E., Smolonska, J., van den Berge, M., Wijmenga, C., Zanen, P., Luinge, M. A., Platteel, M., Lammers, J-W., Dahlback, M., Tosh, K., Hiemstra, P. S., Sterk, P. J., Spira, A., Vestbo, J., Nordestgaard, B. G., Benn, M., Nielsen, S. F., Dahl, M., Verschuren, W. M., ... Lifelines Cohort Study (2014). Susceptibility to chronic mucus hypersecretion, a genome wide association study. *PLoS ONE*, 9(4), [e91621].  
<https://doi.org/10.1371/journal.pone.0091621>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



# Susceptibility to Chronic Mucus Hypersecretion, a Genome Wide Association Study

Akkelies E. Dijkstra<sup>1,2</sup>, Joanna Smolonska<sup>2,3</sup>, Maarten van den Berge<sup>1,2</sup>, Ciska Wijmenga<sup>3</sup>, Pieter Zanen<sup>4</sup>, Marjan A. Luinge<sup>5</sup>, Mathieu Platteel<sup>3</sup>, Jan-Willem Lammers<sup>4</sup>, Magnus Dahlback<sup>6</sup>, Kerrie Tosh<sup>7</sup>, Pieter S. Hiemstra<sup>8</sup>, Peter J. Sterk<sup>9</sup>, Avi Spira<sup>10</sup>, Jorgen Vestbo<sup>11,12</sup>, Borge G. Nordestgaard<sup>13,14</sup>, Marianne Benn<sup>14</sup>, Sune F. Nielsen<sup>13</sup>, Morten Dahl<sup>15</sup>, W. Monique Verschuren<sup>16</sup>, H. Susan J. Picavet<sup>16</sup>, Henriette A. Smit<sup>17</sup>, Michael Owsijewitsch<sup>18,19</sup>, Hans U. Kauczor<sup>18,19</sup>, Harry J. de Koning<sup>20</sup>, Eva Nizankowska-Mogilnicka<sup>21</sup>, Filip Mejza<sup>21</sup>, Pawel Nastalek<sup>21</sup>, Cleo C. van Diemen<sup>3</sup>, Michael H. Cho<sup>22,23,24</sup>, Edwin K. Silverman<sup>22,23,24</sup>, James D. Crapo<sup>25</sup>, Terri H. Beaty<sup>26</sup>, David A. Lomas<sup>27</sup>, Per Bakke<sup>28</sup>, Amund Gulsvik<sup>28</sup>, Yohan Bossé<sup>29</sup>, M. A. Obeidat<sup>30</sup>, Daan W. Loth<sup>31,32</sup>, Lies Lahousse<sup>31,33</sup>, Fernando Rivadeneira<sup>34,35</sup>, Andre G. Uitterlinden<sup>34,35</sup>, Andre Hofman<sup>31,35</sup>, Bruno H. Stricker<sup>31,32</sup>, Guy G. Brusselle<sup>31,33</sup>, Cornelia M. van Duijn<sup>36</sup>, Uilke Brouwer<sup>2,5</sup>, Gerard H. Koppelman<sup>2,37</sup>, Judith M. Vonk<sup>2,38</sup>, Martijn C. Nawijn<sup>2,5</sup>, Harry J. M. Groen<sup>1</sup>, Wim Timens<sup>5</sup>, H. Marike Boezen<sup>2,38</sup>, Dirkje S. Postma<sup>1,2,\*</sup>, the LifeLines Cohort study<sup>1</sup>

1 University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands, 2 University of Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, the Netherlands, 3 University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands, 4 University of Utrecht, University Medical Center Utrecht, Department of Pulmonology, Utrecht, the Netherlands, 5 University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands, 6 AstraZeneca, R&D Mölndal, Mölndal, Sweden, 7 AstraZeneca, Alderley Park, Macclesfield, Cheshire, United Kingdom, 8 Department of Pulmonology, Leiden University Medical Center, Leiden, the Netherlands, 9 Department of Respiratory Medicine, University of Amsterdam, Amsterdam, the Netherlands, 10 Department of Medicine, Section of Computational Biomedicine, Boston University Medical Center, Boston, Massachusetts, United States of America, 11 Odense University Hospital, Department of Respiratory Medicine J, and University of Southern Denmark, Clinical Institute, Odense, Denmark, 12 Respiratory and Allergy Research Group, Manchester Academic Health Sciences Centre, University Hospital South Manchester NHS Foundation Trust, Manchester, United Kingdom, 13 The Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark, 14 Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark, 15 Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark, 16 Center for Prevention and Health Services Research, National Institute for Public Health and the Environment, Bilthoven, the Netherlands, 17 Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, the Netherlands, 18 Department of Diagnostic and Interventional Radiology, University Hospital of Heidelberg, Heidelberg, Germany, 19 Translational Lung Research Center (TLRC-H), Member of the German Center for Lung Research (DZL), Heidelberg, Germany, 20 Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands, 21 Department of Respiratory Diseases, Jagiellonian University Medical College, Krakow, Poland, 22 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, 23 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, 24 Harvard Medical School, Boston, Massachusetts, United States of America, 25 Division of Pulmonary and Critical Care Medicine, National Jewish Health, Denver, Colorado, United States of America, 26 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, 27 Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom, 28 Department of Thoracic Medicine, Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway, 29 Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec City, Canada, 30 Division of Respiratory, Department of Medicine, James Hogg Research Centre, St Paul's Hospital, University of British Columbia, Vancouver, BC, Canada, 31 Dept Epidemiology, Erasmus MC, Rotterdam, the Netherlands, 32 The Netherlands Healthcare Inspectorate, The Hague, the Netherlands, 33 Department of respiratory medicine, University Hospital Ghent, Ghent, Belgium, 34 Department of Internal medicine, Erasmus MC, Rotterdam, the Netherlands, 35 Netherlands Consortium for Healthy Aging (NCHA), Rotterdam, the Netherlands, 36 Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, University Medical Center, Rotterdam, the Netherlands, 37 University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Department of Pediatric Pulmonology and Pediatric Allergy, Groningen, the Netherlands, 38 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands

## Abstract

**Background:** Chronic mucus hypersecretion (CMH) is associated with an increased frequency of respiratory infections, excess lung function decline, and increased hospitalisation and mortality rates in the general population. It is associated with smoking, but it is unknown why only a minority of smokers develops CMH. A plausible explanation for this phenomenon is a predisposing genetic constitution. Therefore, we performed a genome wide association (GWA) study of CMH in Caucasian populations.

**Methods:** GWA analysis was performed in the NELSON-study using the Illumina 610 array, followed by replication and meta-analysis in 11 additional cohorts. In total 2,704 subjects with, and 7,624 subjects without CMH were included, all current or former heavy smokers ( $\geq 20$  pack-years). Additional studies were performed to test the functional relevance of the most significant single nucleotide polymorphism (SNP).

**Results:** A strong association with CMH, consistent across all cohorts, was observed with rs6577641 ( $p = 4.25 \times 10^{-6}$ , OR = 1.17), located in intron 9 of the *special AT-rich sequence-binding protein 1 locus (SATB1)* on chromosome 3. The risk allele (G) was associated with higher mRNA expression of *SATB1* ( $4.3 \times 10^{-9}$ ) in lung tissue. Presence of CMH was associated with increased *SATB1* mRNA expression in bronchial biopsies from COPD patients. *SATB1* expression was induced during differentiation of primary human bronchial epithelial cells in culture.

**Conclusions:** Our findings, that SNP rs6577641 is associated with CMH in multiple cohorts and is a *cis*-eQTL for *SATB1*, together with our additional observation that *SATB1* expression increases during epithelial differentiation provide suggestive evidence that *SATB1* is a gene that affects CMH.

**Citation:** Dijkstra AE, Smolonska J, van den Berge M, Wijmenga C, Zanen P, et al. (2014) Susceptibility to Chronic Mucus Hypersecretion, a Genome Wide Association Study. *PLoS ONE* 9(4): e91621. doi:10.1371/journal.pone.0091621

**Editor:** Dominik Hartl, University of Tübingen, Germany

**Received** July 24, 2013; **Accepted** November 21, 2013; **Published** April 8, 2014

**Copyright:** © 2014 Dijkstra et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The COPACETIC study was funded by EU FP7 grant 201379. The NELSON study was supported by 'Zorg Onderzoek Nederland-Medische Wetenschappen (ZONMW),' KWF Kankerbestrijding,' and 'Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen' (RvvZ). The GLUCOLD study was supported by the Netherlands Organization for Scientific Research, Netherlands Asthma Foundation, University of Groningen, Leiden University Medical Center, and GlaxoSmithKline. The LifeLines cohort study was sponsored by the Dutch ministry of Health, Welfare and Sport, the ministry of Economic Affairs, Agriculture and Innovation, the province of Groningen, the European Union (regional development fund), the Northern Netherlands Provinces (SNN), the Netherlands Organisation for Scientific Research (NWO), University Medical Center Groningen (UMCG), University of Groningen, de Nierstichting (the Dutch Kidney Foundation), and the Diabetes Fonds (the Diabetic Foundation). The ECLIPSE study was funded by GSK and NIH grants RO1 HL089897 and by a grant from GlaxoSmithKline. The COPDGene study was funded by NIH grants RO1 HL089856 and by the COPD foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovion. Data sampling for the Norway study was funded by GlaxoSmithKline. The Dutch Lung Foundation supported the cell culture work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Firms occurring more frequently are coded as follows: A=AstraZeneca; B=Chiesi; C=Merck; D=Boehringer Ingelheim; E=GlaxoSmithKline; F=Nycomed; G=TEVA Pharma; H=Roche; I=Siemens; J=Pfizer. M. Dahlback and K. Tosh are employees of A. The University of Groningen: money for D. S. Postma regarding an unrestricted educational grant for research from A and B, an unrestricted grant for W. Timens from C, a grant from the Dutch Asthma Foundation for W. Timens, and grants for G. H. Koppelman and D. S. Postma from the BBMRI-NL, The Dutch Asthma Foundation and the Stichting Astma Bestrijding, and fees for consultancies of D. S. Postma by A, D, B, E, F, and G, and payment for lectures from A, B, and G and for consultancies of H. J. M. Groen by Eli Lilly, J and H. The University of Leiden: money for P. S. Hiemstra regarding an unrestricted educational grant for research from D, Galapagos. The University of Amsterdam: a public-private grant for P. S. Sterk by Innovative Medicine Initiative by European Union and European Federation of Pharmaceutical Industries and Associations. Boston University Medical Center holds patents. A. Spira: money for board membership, consultancy, and stock options from Allegro Diagnostics, Inc. Money for consultancy and for lectures was paid to J. Vestbo by E, F, B, Syntaxis, Bioxydyn, Novartis, A, and D. University Hospital of Heidelberg: money for expert testimony for H.U.Kauczor. H. U. Kauczor: money for board membership from I, for lectures from Bracco, I, D, and Bayer, and for development of educational presentations from I. The Jagiellonian University Medical College of Krakow: a Grant (Z/7PR/00010,201 379) for E. Nizankowska-Mogilnicka. E. Silverman: a grant for Harvard University from E. E. Silverman: consulting fee and support for travel to meetings for the study or other purposes (COPDGene and ECLIPSE) from E and the COPD foundation and money for consultancy and lectures from E, A and C. M. Cho: NIH/NHLBI grants (K12HL089990 and K08 HL097029) for Harvard University and money for consultancy from C. Erasmus Medical Center Rotterdam: a grant for H. de Koning from ZonMw, KWF, Kankerbestrijding, Stichting Centraal Fonds Reserves van Voormalig Vrijwillige Ziekenfondsverzekeringen; H provided a grant for the performance of proteomics research, and I provided 4-digital workstations for LungCare for the performance of 3Dmeasurements. Johns Hopkins Bloomberg School of Public Health in Baltimore: a NIH grant for J. Crapo. The National Jewish Health Center in Denver: a grant for T. Beaty from the National Institutes of Health and National Institute Heart, Lung & Blood Disease. T. Beaty: money for travel support by the COPD foundation. University of Cambridge: a grant for D. Lomas from E. D. Lomas: money for board membership, consultancy, lectures, and travel accommodations from E, Talecris/Grifols, and D. P. Bakke: money for lectures from E, A, and J. L. Lahousse: a travel award for the ATS (2012) and a short-term Research Fellowship for transport to Rotterdam. Part of the GLUCOLD study was supported by GlaxoSmithKline, the ECLIPSE study was partly funded by GSK, and the COPDGene study was partly funded by the COPD foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovion. Data sampling for the Norway study was funded by GlaxoSmithKline. There are no further patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors. Patent nr. subjectdate 1US 2013/0023437Diagnostic for lung disorders using class prediction 2012-06-15 2US 2012/0322673Isolation of nucleic acid from mouth epithelial cells 2012-03-23 3US 2012/0190567Diagnostic and prognostic methods for lung disorders using gene expression profiles from nose epithelial cells 2011-12-12 4US 2011/0217717Diagnostic and prognostic methods for lung disorders using gene expression profiles from nose epithelial cells2010-11-05 5US 2011/0190150Isolation of nucleic acid from mouth epithelial cells 2010-09-17 6US 2012/0041686Diagnostic for lung disorders using class prediction 2010-08-26 7US 2010/0055689Multifactorial methods for detecting lung disorders2009-03-30 8US 2009/0311692Isolation of nucleic acid from mouth epithelial cells 2009-01-09 9US 2009/0186951Identification of novel pathways for drug development for lung disease 2008-09-19 10US 2009/0061454Diagnostic and prognostic methods for lung disorders using gene expression profiles from nose epithelial cells 2007-03-08 11 US 2010/0035244Diagnostic for lung disorders using class prediction2006-04-14 12US 2006/0154278Detection methods for disorders of the lung2005-12-06 13US 2007/0148650Isolation of nucleic acid from mouth epithelial cells2004-11-12.

\* E-mail: d.s.postma@umcg.nl

¶ Membership of the LifeLines Cohort study is provided in the Acknowledgments.

## Introduction

The secretion of mucus is a natural part of the airway defense against inhaled noxious particles and substances. Chronic mucus hypersecretion (CMH) is a condition of overproduction of mucus and defined as the presence of sputum production during at least three months in two consecutive years without any explaining origin whereas airway obstruction is not a prerequisite [1]. Smoking is a risk factor for CMH, i.e. the prevalence of CMH in the general population is reported to be 7.4% in current smokers, 3.7% in ex-smokers and 2.4% in never smokers [2]. CMH is the key presenting symptom in chronic bronchitis, one of the three main sub-groups of chronic obstructive pulmonary disease (COPD), a complex disease characterized by the presence of incompletely reversible and generally progressive airflow limitation [3]. Moreover, CMH is a risk factor for the development of COPD [4,5].

Worldwide, COPD affected 65 million people in 2004 and more than 3 million people died of COPD in 2005, representing 5% of all deaths. It is predicted that COPD will be the third

leading cause of death worldwide in 2030 [6]. COPD markedly reduces quality of life and is responsible for high healthcare costs. For instance, the combined (direct and indirect) yearly costs of COPD and asthma in the United States of America were projected at \$68 billion in 2008 [7]. CMH is not only associated with COPD but also with an increased duration and frequency of respiratory infections, excess decline in forced expiratory volume in 1 second (FEV<sub>1</sub>) and increased hospitalization and mortality rates in the general population [4,5,8,9].

It is not known why only a minority of all smokers develops CMH, yet a plausible explanation is the presence of a genetic predisposition for CMH, as evidenced by familial aggregation of mucus overproduction and higher prevalence of CMH in monozygotic than in dizygotic twins [10–12]. Little is known about the identity of the genes that predispose to CMH. One publication suggested that *CTLA4* is associated with chronic bronchitis in COPD [13].

The aim of our study was to identify genetic factors for CMH, thereby obtaining a better insight into the origins of this disorder.

**Table 1.** Questions used to define chronic mucus hypersecretion in the corresponding cohorts.

Cohort	Question
NELSON [24]	Do you expectorate sputum on the majority of days more than 3 months a year, even when you do not have a cold?
Rotterdam [25,26]	Do you expectorate sputum on the majority of days during $\geq 3$ months during the last 2 years?
LifeLines [27]	Do you usually expectorate sputum during day or night in winter? If yes: Do you expectorate sputum on the majority of days $>3$ months a year?
Vlagentwede- Vlaardingen [28,29]	Do you expectorate sputum on the majority of days $>3$ months a year?
Doetinchem [30]	Do you expectorate sputum during winter, day and night, each day for 3 months?
Poland [31,32]	Do you usually bring up phlegm from your chest, or do you usually have phlegm in your chest that is difficult to bring up when you don't have a cold? If yes: Are there months in which you have this phlegm on most days? If yes: Do you bring up this phlegm on most days for as much as three months each year? A positive answer to all (3) questions identifies CMH.
Heidelberg [33]	Do you expectorate sputum on the majority of days $>3$ months a year?
GLUCOLD [15]	Do you expectorate sputum immediately after getting up on the majority of days in winter $>3$ months a year?
Rucphen [29]	Do you expectorate sputum during day or night in winter? If yes: Do you have expectoration on the majority of days $>3$ months a year?
ECLIPSE [34]	Do you usually bring up phlegm from your chest on getting up, first thing in the morning, during the rest of the day or at night, on most days for 3 consecutive months or more during the year?
COPDGene [35]	Do you usually bring up phlegm from your chest on getting up, first thing in the morning, during the rest of the day or at night, on most days for 3 consecutive months or more during the year?
Norway [36,37]	Do you usually bring up phlegm from your chest on getting up, first thing in the morning, during the rest of the day or at night, on most days for 3 consecutive months or more during the year?

doi:10.1371/journal.pone.0091621.t001

## Materials and Methods

### Ethics Statement

The Dutch ministry of health and the Medical Ethics Committee of the hospital approved the study protocol for all Dutch centers. Ethics approval and written informed consent was obtained from all participants in all studies participating. For detailed information, see Supplement S1.

### Subjects and genotyping

We performed GWA studies in participants of the NELSON-study ( $n = 3,729$ ), a male population-based lung cancer screening study investigating heavy smokers ( $\geq 20$  pack-years) [24].

Replication of SNPs with  $p \leq 10^{-4}$  was attempted in six cohorts participating in 'COPD Pathology: Addressing Critical gaps, Early Treatment & diagnosis and Innovative Concepts' (COPACETIC) and in five non-COPACETIC cohorts. Caucasian subjects with  $\geq 20$  pack-years smoking with genotype-, spirometric- and demographic data were included.

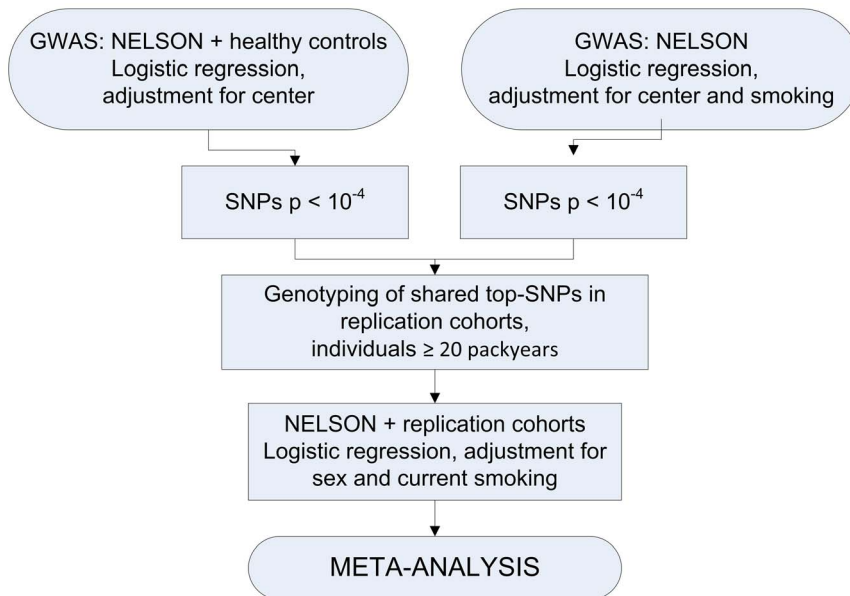
An overview of the CMH definitions used in this study is presented in Table 1. A brief description of the included cohorts and details according to the period of data collection, type of population, genotyping platforms and genetic imputation software are presented in Table 2.

**Table 2.** Overview of populations.

Study	Data Collection	Type of population	Genotyping platform	Imputation software
NELSON	2004/2005	general population	Illumina Quad 610	NA
GLUCOLD	2005/2006	COPD case	Illumina Veracode	NA
Vlagentwede Vlaardingen	1989/1990	general population	Illumina Veracode	NA
Doetinchem	1998/2002	general population	Illumina Veracode	NA
Poland	2005/2006	general population	Illumina Veracode	NA
Heidelberg	2004/2005	general population	Illumina Veracode	NA
Rucphen	2002	Family based COPD on a doctor diagnosis	Illumina Veracode	NA
Rotterdam	2002/2008	general population	Illumina 550K	MaCH
LifeLines	2008/2010	general population	Illumina Human CytoSNP-12	BEAGLE v3.1.0
COPDGene	2008/2009	COPD case/control (stage I–IV)	Illumina Human Omni1-Quad	MaCH
ECLIPSE	2005/2007	COPD case/control (stage II–IV)	Illumina Human HAP 550 V3	MaCH
Norway	2003/2005	COPD case/control (stage II–IV)	Illumina Human HAP 550 V1, V3, and DUO	MaCH

Populations and corresponding period of data collection, type of population, genotyping platform and soft-ware used for imputation. NA = not applicable.

doi:10.1371/journal.pone.0091621.t002



**Figure 1. Study design.** We performed GWA studies in the NELSON cohort and in additional healthy controls. CMH was analyzed using logistic regression with adjustment for center (Groningen and Utrecht). Since current smoking can affect the presence of CMH, we additionally performed the GWAS in the NELSON cohort correcting for center and smoking. SNPs with a  $p$ -value  $< 10^{-4}$  present in both GWA studies were selected for replication. To test for generalizability of associations with CMH in other populations, we compared our results with data in CMH-cases and controls with a smoking history of  $\geq 20$  pack-years with eleven replication populations using logistic regression with adjustment for sex and current smoking. Finally, we performed a meta-analysis on shared SNPs across the NELSON identification population and the 11 replication populations. doi:10.1371/journal.pone.0091621.g001

## Strategy

We searched for SNPs associated with CMH by using a two-stage strategy followed by a replication stage and meta-analysis (Figure 1).

## Statistical analysis

General characteristics of CMH-cases and controls were compared using Student's  $t$ - and Mann-Whitney- $U$  tests for continuous variables and  $\chi^2$  tests for dichotomous variables with SPSS 20.0. Sample and SNP quality control (QC), regression- and meta-analysis were performed with PLINK 1.07 [25]. QC criteria are described in Supplement S1.

Logistic regression analysis under an additive model was used to identify SNPs associated with CMH. SNPs with a  $p$ -value  $< 10^{-4}$  were included for replication. When two SNPs were in strong linkage disequilibrium ( $r^2 \geq 0.8$ ), the SNP with the lowest  $p$ -value was further analyzed.

SNPs in COPACETIC cohorts and in LifeLines were analyzed using logistic regression with adjustment for sex and smoking (ex-/current smoking). In LifeLines, imputed SNPs with an info-score  $< 0.3$  (imputation quality score) were removed. SNPs in non-COPACETIC cohorts were analyzed by the cohort investigators using the same model.

Meta-analysis was performed on SNPs across NELSON and the 11 replication cohorts. The Cochran's  $Q$  test was used to test for heterogeneity in the meta-analysis.

We performed multivariate logistic regression analysis, adjusted for pack-years and lung function, to associate CMH with the risk allele of rs6577641 in the identification cohort.

## Functional relevance of *SATB1* and rs6577641, our highest ranked-SNP

We performed 4 functional studies with the identified top-SNP. Details on their methods are given in Supplement S1.

We assessed:

- 1) whether rs6577641 is an eQTL, by analyzing the association of *SATB1* expression levels with rs6577641 genotypes in lung tissue from three independent cohorts recruited from Laval University, University of British Columbia, and University of Groningen as described previously [14];
- 2) CMH-associated mRNA expression in airway wall biopsies from 77 COPD participants in the GLUCOLD-study [15];
- 3) the association of homozygous genotypes for rs6577641 with a) immunohistochemical staining (IHC) for *SATB1* and b) the fraction of mucus positivity on bronchial tissue explanted from COPD or lung cancer subjects that underwent lung surgery;
- 4) *SATB1* expression levels during mucociliary differentiation of primary bronchial epithelial cells cultured at air-liquid interface [26].

## Results

### Populations

Characteristics of the identification and replication populations are presented in Table 3. Subjects with CMH were more often current smokers and had worse lung function, except for populations including subjects with COPD only.

### Identification analysis

After QC, 492,700 SNPs and 2,512 individuals (717 CMH cases, 1,795 controls) from the NELSON study remained. Logistic regression analysis was performed including these individuals

**Table 3.** Demographic and clinical characteristics of CMH-cases and -controls with  $\geq 20$  pack-years, present in the meta-analysis.

Population	CMH	N	Population %	Female, %	Age, yrs (SD)	Pack-years (range)	Current smoking, %	FEV1 %, pred. (SD)	FEV1/FVC, % (SD)
NELSON	Control	1,795	71.5	0	60.2 (5.3)	34 (21–156)	47.5	100.3 (17.2)	72.9 (8.7)
NELSON	Case	717	28.5	0	60.4 (5.6)	39 (21–140)	74.2	93.5 (20.0)	69.2 (11.0)
Rotterdam	Control	1,043	84.1	46.1	68.0 (9.3)	45 (20–149)	40.1	92.4 (23.5) <sup>#</sup>	72.8 (8.7) <sup>#</sup>
Rotterdam	Case	197	15.9	43.7	72.0 (8.4)	40 (20–168)	45.2	85.0 (26.9) <sup>#</sup>	68.0 (11.1) <sup>#</sup>
LifeLines	Control	1,431	88.1	80.1	52.9 (9.2)	27 (20–100)	56.4	98.2 (15.6)	72.4 (8.2)
LifeLines	Case	193	11.5	46.9	53.2 (9.9)	29 (20–97)	75.4	90.5 (18.0)	68.3 (11.3)
Vlagtwedde-Vlaardingen*	Control	234	82.4	27.4	52.9 (10.1)	29 (20–128)	51.7	94.5 (12.1)	76.6 (4.5)
Vlagtwedde-Vlaardingen*	Case	50	17.6	18	53.4 (10.5)	33 (22–83)	68	86.7 (18.6)	71.0 (8.9)
Doetinchem	Control	250	80.6	37.2	54.7 (8.8)	30 (20–90)	55.6	94.8 (17.6)	71.5 (9.9)
Doetinchem	Case	60	19.4	36.7	56.4 (7.7)	33 (20–72)	68.3	89.1 (19.6)	69.3 (11.4)
Poland	Control	97	85.1	22.7	56.7 (10.5)	30 (20–116)	52.6	96.4 (21.4)	72.5 (0.5)
Poland	Case	17	14.9	11.8	55.8 (9.4)	35 (22–86)	82.4	93.5 (24.0)	69.2 (13.1)
Heidelberg	Control	608	84.2	35.7	58.1 (5.2)	37 (23–138)	54.3	96.4 (17.6)	78.9 (9.7)
Heidelberg	Case	114	15.8	29.8	58.0 (5.2)	37 (23–91)	91.2	86.2 (21.5)	75.3 (10.6)
GLUCOLD**	Control	48	55.2	8.3	62.6 (7.6)	46 (21–182)	62.5	63.4 (9.8)	50.4 (9.1)
GLUCOLD**	Case	39	44.8	20.5	59.6 (7.4)	40 (22–83)	61.5	63.9 (8.8)	53.1 (7.8)
Rucphen**	Control	28	53.8	46.4	66.5 (7.9)	42 (21–120)	57.1	74.5 (15.7)	57.2 (7.8)
Rucphen**	Case	24	46.2	41.7	62.2 (10.5)	43 (21–100)	70.8	70.2 (21.6)	53.1 (9.7)
ECLIPSE**	Control	961	62	37.5	64.1 (6.7)	53 (21–205)	28.1	48.0 (15.7)	44.5 (11.3)
ECLIPSE**	Case	590	38	24.1	62.9 (7.4)	54 (22–220)	47.5	46.2 (15.5)	44.3 (11.7)
COPDGene	Control	628	71.8	53.5	63.1 (8.6)	50 (21–173)	28.2	75.0 (28.3)	63.7 (17.6)
COPDGene	Case	247	28.2	40.5	61.9 (8.4)	54 (21–237)	50.2	60.4 (27.4)	54.6 (17.9)
Norway	Control	501	52.4	44.9	61.5 (10.3)	34 (20–130)	46.9	71.7 (24.2)	64.6 (15.7)
Norway	Case	456	47.6	20.4	64.1 (10.1)	39 (20–119)	59	56.5 (24.4)	55.0 (17.3)

CMH = Chronic mucus hypersecretion;

\*lung function is based on FEV<sub>1</sub>/FVC;

\*\*all individuals in this cohort have COPD;

<sup>#</sup>based on lung function of 700 subjects who returned for follow-up study 4 years later.

doi:10.1371/journal.pone.0091621.t003

supplemented with 590 additional healthy controls, adjusting for center. The QQ-plot provided no evidence of population stratification ( $\lambda = 1.0185$ ). 77 SNPs were associated with CMH with a p-value  $< 10^{-4}$ . CMH was associated with current smoking in our identification cohort ( $p < 0.001$ ). Therefore, we performed a second GWA adjusting for center and current/ex-smoking (717 CMH-cases, 1,795 controls). The QQ-plot showed no evidence of population stratification ( $\lambda = 1.0056$ ). We observed 64 SNPs with a p-value  $< 10^{-4}$ . Genome wide association for CMH ordered by chromosome is shown in the Manhattan plot. Figure 2 shows QQ-plots (A, C) and genome wide association signals for CMH ordered by chromosome (Manhattan-plots, B and D) of these sequential analyses. We identified 36 SNPs associated with CMH with a p-value  $< 10^{-4}$  in both analyses Table 4. Of these, 32 SNPs were included for replication and 4 SNPs were removed because they were in strong linkage disequilibrium ( $r^2 > 0.8$ ) with another associated SNP.

### Replication of associated SNPs

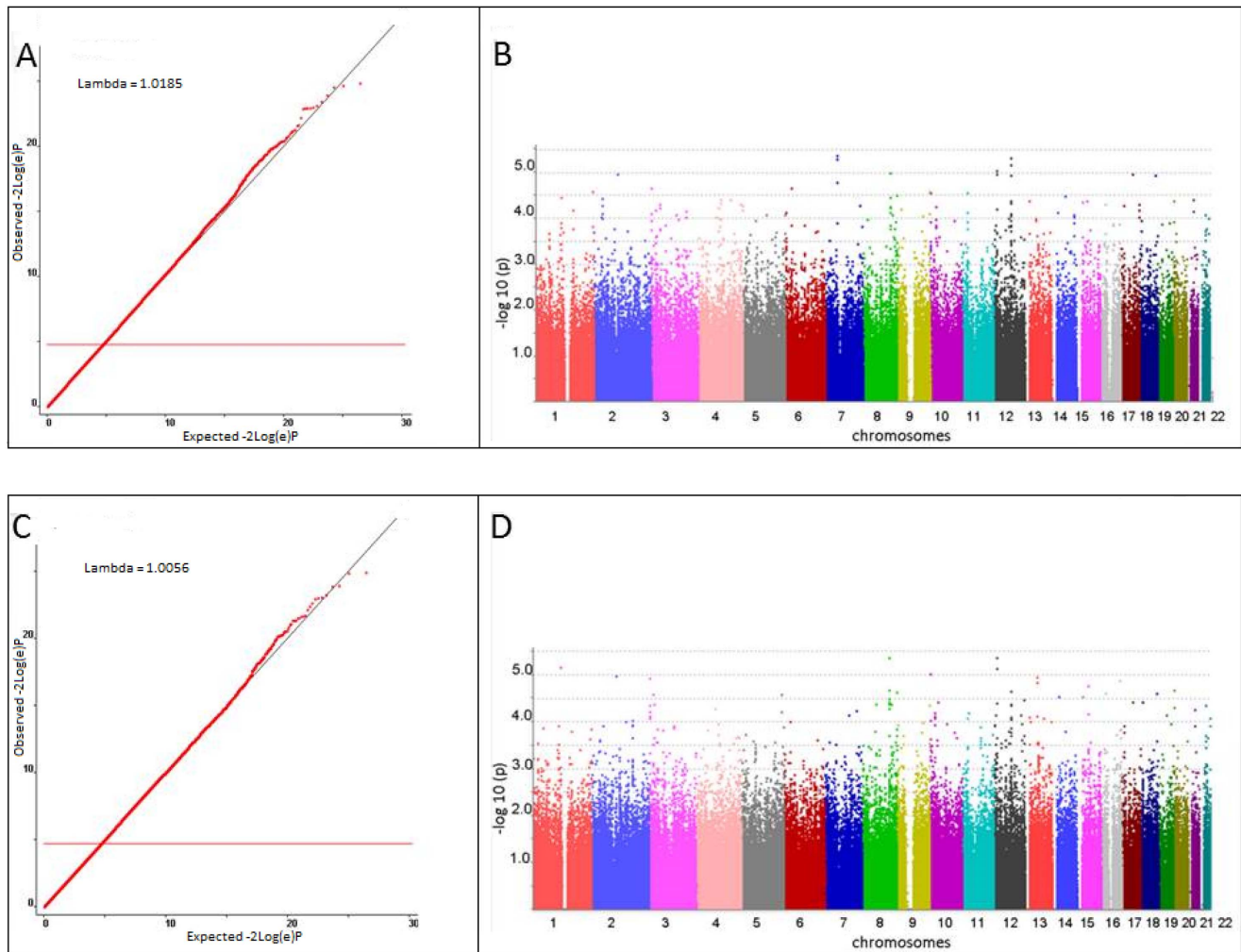
Genotyping of SNP rs4775569 failed in the COPACETIC populations, and was removed for further analysis. CMH-associated top-SNPs for each cohort are presented in Table 5,

with a complete overview in Table 6. When applying Bonferroni correction in the meta-analysis ( $p = 1.61 \times 10^{-3}$  for 31 SNPs), we found a strong association with one SNP:

- rs6577641, a SNP located on chromosome 3 in intron 9 of the *special AT-rich sequence-binding protein 1 locus (SATB1)* (combined p-value =  $4.25 \times 10^{-6}$ , OR = 1.17; 1.10–1.26).

The *SATB1* SNP rs6577641 had the lowest p-value for association with CMH in the meta-analysis. Figure 3 shows the forest plot of rs6577641 in the identification and replication cohorts and meta-analysis.

We assessed the percentage of subjects with CMH in each genotyping group for rs6577641 in NELSON-total and stratified for current and ex smokers (Figure 4). Multivariate logistic regression analysis, corrected for pack-years and FEV<sub>1</sub>% predicted, showed that CMH was significantly associated with the number of G-alleles in the 1,385 current smokers (reference = AA: heterozygous mutant (AG)  $p = 0.001$ ; OR = 1.50, homozygous mutant (GG)  $p = 0.001$ ; OR = 1.80) but not in 1,127 ex-smokers (reference = AA: heterozygous mutant (AG)  $p = 0.380$ ; OR = 1.18, homozygous mutant (GG)  $p = 0.143$ ; OR = 1.42).



**Figure 2. Quantile-quantile plot and Manhattan plot of GWA results for association of SNPs with CMH in NELSON amplified with bloodbank controls and corrected for center (A and B).** Quantile-quantile plot and Manhattan plot of GWA results for association of SNPs with CMH in NELSON, corrected for center and smoking habits (C and D). doi:10.1371/journal.pone.0091621.g002

### Functional relevance of *SATB1* and rs6577641

#### 1) Transcriptional regulation of *SATB1* mRNA expression

We analyzed the association of *SATB1* expression levels in lung tissue with rs6577641 genotype in 3 independent data sets of the Universities of Groningen, Laval and UBC [14]. A *cis*-acting effect of rs6577641 on *SATB1* expression was identified and present in all three datasets ( $n = 1,095$ ), with the same direction of effect across all three *SATB1* probes on the array. The (susceptibility) G allele increased expression, the (protective) A allele reduced expression ( $p = 4.3 \times 10^{-9}$ ) in the meta-analysis across the three datasets and across all three *SATB1* probes measured (Table 7).

#### 2) *SATB1* mRNA expression and CMH

We compared *SATB1* expression in baseline airway wall biopsies of COPD patients with ( $n = 38$ ) and without ( $n = 39$ ) CMH in GLUCOLD [15]. CMH was significantly associated with *SATB1* expression levels (corrected for ex-/current smoking;  $p = 0.0045$ ; Figure 5). After stratification, the same direction of effect was present in ex- and current smokers. However, this association reached statistical significance in current smokers ( $p = 0.021$ ) and not in ex- smokers ( $p = 0.132$ ), probably due to a difference in power as 46 subjects were current smokers versus 33 ex-smokers.

#### 3) Genotype related protein expression and mucus positivity in bronchial epithelium

*SATB1* protein expression has previously been observed in IHS analysis of bronchial epithelial cells [16]. Therefore, we stained *SATB1* on paraffin embedded lung tissue biopsies of individuals from the Groningen population contributing to the eQTL analysis. We observed clear nuclear staining for *SATB1* in bronchial epithelial cells. No significant difference for % of strong positive, positive and weak positive cells was observed between the protective (AA,  $n = 9$ ) and risk (GG,  $n = 14$ ) rs6577641 genotypes ( $11.8\% \pm 5.8$  versus  $12.7\% \pm 6.9$ ,  $p = 0.74$ ).

We determined whether the fraction of mucus positive bronchial epithelium was different in subjects with different homozygous rs6577641 genotypes and performed PAS-staining on tissue biopsies from the same cohort. We observed no significant difference between individuals with the homozygous protective (AA,  $n = 10$ ) and risk (GG,  $n = 7$ ) alleles ( $19.7\% \pm 11.9$  versus  $14.3\% \pm 9.6$ ,  $p = 0.34$ ).

#### 4) *SATB1* expression levels during bronchial epithelial cell mucociliary differentiation

We investigated whether *SATB1* expression was induced during mucociliary differentiation of primary human bronchial epithelial

**Table 4.** SNPs associated with CMH with a p-value <math>10^{-4}</math>, present in GWAS-I and in GWAS-II, in the NELSON identification cohort.

Chromosome	SNP	Base pair position	p-value GWAS I	p-value GWAS II
2	rs6735868	103582093	$1.11 \times 10^{-05}$	$1.08 \times 10^{-05}$
3	rs1387089	1940922	$7.94 \times 10^{-05}$	$4.56 \times 10^{-05}$
3	rs1488757	1981567	$2.17 \times 10^{-05}$	$1.16 \times 10^{-05}$
3	rs6577641	18397849	$6.83 \times 10^{-05}$	$2.57 \times 10^{-05}$
4	rs4306981	79924121	$9.74 \times 10^{-05}$	$5.18 \times 10^{-05}$
8	rs4242562	115475287	$7.66 \times 10^{-05}$	$5.13 \times 10^{-05}$
8	rs7836298	115504434	$1.03 \times 10^{-05}$	$4.37 \times 10^{-06}$
8	rs7823554*	115553109	$6.05 \times 10^{-05}$	$5.22 \times 10^{-05}$
8	rs7836963*	115568426	$5.52 \times 10^{-05}$	$4.24 \times 10^{-05}$
8	rs16886291	115711436	$3.54 \times 10^{-05}$	$2.09 \times 10^{-05}$
8	rs10098746	125838127	$8.47 \times 10^{-05}$	$4.34 \times 10^{-05}$
8	rs7831595	144974963	$3.08 \times 10^{-05}$	$2.32 \times 10^{-05}$
9	rs4842047	138816796	$2.63 \times 10^{-05}$	$4.51 \times 10^{-05}$
10	rs943189	22842590	$5.57 \times 10^{-05}$	$6.33 \times 10^{-05}$
11	rs11026531	22379184	$2.76 \times 10^{-05}$	$8.55 \times 10^{-05}$
12	rs1894307*	12005720	$9.04 \times 10^{-06}$	$7.18 \times 10^{-06}$
12	rs2255953	12010736	$1.13 \times 10^{-05}$	$4.33 \times 10^{-06}$
12	rs2855708	12013572	$6.47 \times 10^{-05}$	$3.97 \times 10^{-05}$
12	rs10879509*	73242131	$6.98 \times 10^{-06}$	$4.44 \times 10^{-05}$
12	rs4760851	73284781	$4.85 \times 10^{-06}$	$2.29 \times 10^{-05}$
12	rs952394	73441110	$4.18 \times 10^{-05}$	$4.22 \times 10^{-05}$
12	rs12822199	75458164	$4.82 \times 10^{-05}$	$8.58 \times 10^{-05}$
12	rs1379963	75493882	$1.18 \times 10^{-05}$	$2.20 \times 10^{-05}$
12	rs1795669	76273692	$8.01 \times 10^{-05}$	$7.86 \times 10^{-05}$
13	rs9578362	21882381	$4.28 \times 10^{-05}$	$7.99 \times 10^{-05}$
13	rs1211304	50381016	$9.96 \times 10^{-05}$	$1.12 \times 10^{-05}$
14	rs992745	27810095	$7.67 \times 10^{-05}$	$2.99 \times 10^{-05}$
15	rs754661	26934277	$4.54 \times 10^{-05}$	$2.88 \times 10^{-05}$
15	rs4775569	46850317	$4.20 \times 10^{-05}$	$1.72 \times 10^{-05}$
16	rs13333521	19904082	$5.08 \times 10^{-05}$	$2.50 \times 10^{-05}$
17	rs11652469	49565797	$1.13 \times 10^{-05}$	$3.80 \times 10^{-05}$
18	rs8086262	69227590	$1.15 \times 10^{-05}$	$2.53 \times 10^{-05}$
20	rs4815628	3891896	$4.17 \times 10^{-05}$	$2.15 \times 10^{-05}$
21	rs2032257	27774870	$3.97 \times 10^{-05}$	$5.39 \times 10^{-05}$
22	rs1009147	30088257	$8.41 \times 10^{-05}$	$4.51 \times 10^{-05}$
22	rs1005239	47687170	$9.86 \times 10^{-05}$	$8.67 \times 10^{-05}$

\*SNP not selected for replication because of strong linkage disequilibrium with another SNP.  
doi:10.1371/journal.pone.0091621.t004

(HBE) cells *in vitro* and compared *SATB1* mRNA expression levels at different time points of an air-liquid interface (ALI) culture for up to 45 days. ALI culture of HBE cells induced mucociliary differentiation, as confirmed by induction of expression of *FOXJ1*, a marker for ciliated cells (19) and *MUC5AC*, a marker of goblet cells. *SATB1* expression was induced over time (Figure 6), with an approximately 8-fold increased expression from the start to the end of the 45-day ALI culture period.

## Discussion

Since not every ex- or current heavy smoker suffers from chronic mucus hypersecretion (CMH), we aimed to identify

genetic variants conferring susceptibility to CMH. Therefore, we performed the first GWA study on CMH, the key presenting symptom in chronic bronchitis. CMH was associated with 36 SNPs at the  $p < 10^{-4}$  significance level in the identification cohort. In the meta-analysis combining our identification and replication cohorts, strong association was observed with rs6577641, a SNP located on chromosome 3 in intron 9 of *SATB1*. Although the association of rs6577641 with CMH did not reach conventional genome-wide significance, its effect was in the same direction and was significant ( $4.25 \times 10^{-6}$ ) at nominal levels ( $1.61 \times 10^{-3}$ ) across eleven study populations, showing the robustness of this finding. The detected odds ratio for this SNP suggests an additional risk of



**Table 5.** Meta-analysis of top SNPs associated with CMH in replication cohorts, in identification and replication cohorts and corresponding direction of effect in all cohorts and associated feature and gene(s).

Chr	SNP	Base pair position	Minor allele	MAF	Meta-analysis across replication cohorts			Meta-analysis across identification and replication cohorts			Close(st) genes
					p-value	OR	Direction of effect per cohort*	p-value	OR	Q	
3	rs6577641	18397849	G	0.400	5.01E-03	1.12	++++++0+0+	4.25 E-06	1.17	6.20E-01	<i>SATB1</i> #
3	rs1488757	1981567	G	0.109	2.34E-01	0.92	-00-----0	1.10E-03	0.83	1.55E-01	<i>LOC727810 and CNTN4</i>
12	rs2855708	12013572	G	0.273	2.18E-01	1.06	+0+0-++++0	1.20E-03	1.13	1.76E-01	<i>ETV6</i> #
14	rs992745	27810095	G	0.234	2.94E-01	0.95	-+-----++	2.74E-03	0.89	4.59E-02	<i>LOC7288755</i> #
4	rs4306981	79924121	G	0.307	3.37E-01	1.04	++0++++0+	1.38E-03	1.12	5.19E-02	<i>PAQR3 and ARD1B</i>
12	rs1795669	76273692	A	0.059	2.83E-01	1.09	+++++-----++	2.90E-03	1.22	1.77E-01	<i>LOC100130336 and LOC100131830</i>
9	rs4842047	137956617	A	0.303	3.88E-01	0.96	-0-XX+X0-00	3.44E-03	0.89	3.03E-01	<i>CAMPSAP1 and UBAC1</i>
13	rs95788362	21882381	A	0.402	8.05E-01	1.01	++++0--00	3.61E-03	0.91	2.88E-02	<i>LOC6500794 and GRK6PS</i>
12	rs2255953	12010736	G	0.212	5.31E-01	0.97	+X---+0++0	5.12E-03	1.13	4.54E-02	<i>ETV6</i> #
15	rs754661	26934277	G	0.405	5.45E-01	0.96	-00X-+++0+	6.29E-03	0.91	1.08E-01	<i>GABRB3</i> #
8	rs16886291	115711436	A	0.127	5.01E-03	1.12	---+--+00+	5.41E-03	0.86	1.55E-01	<i>hCG_1644355 and TRPS1</i>

MAF = minor allele frequency in NELSON;

\*Direction of effect per cohort: each sign reflects one cohort, direction of effect is presented by: += (OR&gt;1.05), -= (OR&lt;0.95), 0 = (0.95&lt;OR&lt;1.05) and x = (missing result); cohorts are presented in the same order as in Table 2; OR is odds ratio; Q is p-value for heterogeneity;

# means corresponding SNP is located in an intron in this gene.

doi:10.1371/journal.pone.0091621.t005

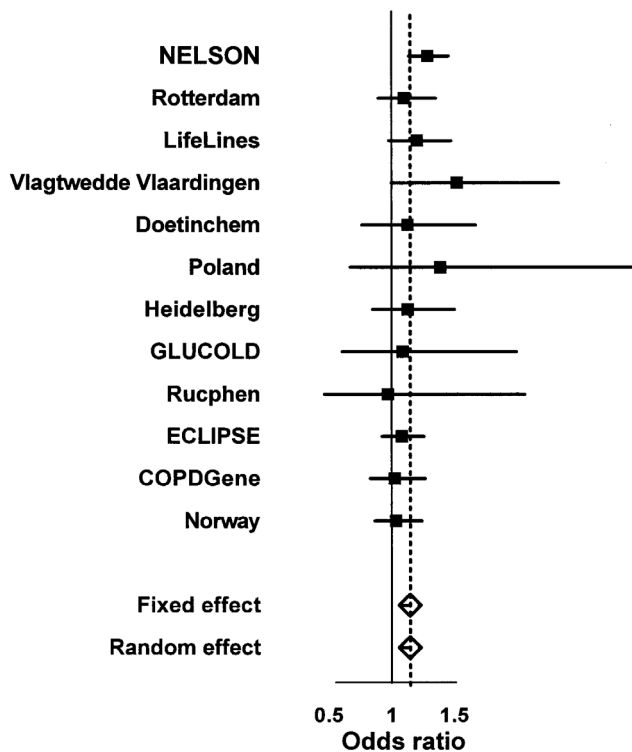
**Table 6.** Meta-analysis of top SNPs associated with CMH across replication cohorts and across identification and replication cohorts, corrected for smoking and sex.

Chromosome	Base pair position	SNP	Minor allele	Meta-analysis across replication cohorts				Meta-analysis across identification and replication cohorts				Close(st) gene(s)			
				p-value	OR	Q	N	p-value	OR	Q	N				
3	18397849	rs6577641	G	5.01E-03	1.121	9.19E-01	12	4.25E-06	1.173	6.20E-01	12	4.25E-06	1.173	6.20E-01	<i>SATB1*</i>
18	69227590	rs8086262	G	2.16E-02	1.107	1.00E-02	11	8.91E-02	1.129	2.60E-03	11	8.91E-02	1.129	2.60E-03	<i>LOC100132647 and CBLN2</i>
8	115475287	rs4242562	C	5.04E-02	1.175	4.20E-01	12	6.15E-01	1.073	1.40E-03	12	6.15E-01	1.073	1.40E-03	<i>hCG_1644355 and TRPS1</i>
8	125838127	rs10098746	A	5.74E-02	0.906	1.76E-01	9	6.22E-01	0.954	2.00E-04	9	6.22E-01	0.954	2.00E-04	<i>MTSS1 and LOC100130448</i>
12	75493882	rs1379963	A	8.87E-02	1.089	6.08E-01	11	4.57E-01	1.063	2.10E-03	11	4.57E-01	1.063	2.10E-03	<i>KCNC2*</i>
12	75458164	rs12822199	G	9.66E-02	1.093	5.30E-01	12	8.50E-01	1.016	4.70E-03	12	8.50E-01	1.016	4.70E-03	<i>KCNC2*</i>
13	50381016	rs1211304	A	1.37E-01	1.099	6.86E-01	11	8.70E-01	0.983	2.50E-03	11	8.70E-01	0.983	2.50E-03	<i>KPNA3 and LOC220429</i>
12	12013572	rs2855708	G	2.18E-01	1.057	7.02E-01	12	1.20E-03	1.132	1.76E-01	12	1.20E-03	1.132	1.76E-01	<i>ETV6*</i>
3	1981567	rs1488757	G	2.34E-01	0.923	8.44E-01	12	1.10E-03	0.828	1.55E-01	12	1.10E-03	0.828	1.55E-01	<i>LOC727810 and CNTN4</i>
12	76273692	rs1795669	A	2.83E-01	1.087	7.02E-01	12	2.90E-03	1.218	1.77E-01	12	2.90E-03	1.218	1.77E-01	<i>LOC100130336 and LOC100131830</i>
14	27810095	rs992745	G	2.94E-01	0.952	4.06E-01	12	2.74E-03	0.886	4.59E-02	12	2.74E-03	0.886	4.59E-02	<i>LOC728755*</i>
13	21882381	rs9578362	A	3.06E-01	0.96	2.06E-01	12	3.61E-03	0.905	2.88E-02	12	3.61E-03	0.905	2.88E-02	<i>LOC650794 and GRK6PS</i>
4	79924121	rs4306981	G	3.37E-01	1.043	3.45E-01	12	2.89E-03	1.117	5.19E-02	12	2.89E-03	1.117	5.19E-02	<i>PAQR3 and ARD1B</i>
2	103582093	rs6735868	A	3.66E-01	1.051	8.51E-01	12	1.59E-01	0.936	1.38E-02	12	1.59E-01	0.936	1.38E-02	<i>TMEM182 and LOC728815</i>
9	137956617	rs4842047	A	3.88E-01	0.961	9.99E-01	9	3.44E-03	0.893	3.03E-01	9	3.44E-03	0.893	3.03E-01	<i>CAMSAP1 and UBAC1</i>
17	49565797	rs11652469	G	4.51E-01	0.942	2.39E-01	12	9.77E-01	1.004	3.50E-03	12	9.77E-01	1.004	3.50E-03	<i>FLJ42842 and LOC388401</i>
15	26934277	rs754661	G	5.31E-01	0.974	8.08E-01	11	6.29E-03	0.908	1.08E-01	11	6.29E-03	0.908	1.08E-01	<i>GABRB3*</i>
8	115711436	rs16886291	A	5.45E-01	0.963	8.77E-01	12	6.47E-03	0.864	1.32E-01	12	6.47E-03	0.864	1.32E-01	<i>hCG_1644355 and TRPS1</i>
20	3891896	rs4815628	A	5.90E-01	1.022	2.76E-01	12	4.51E-01	0.954	3.90E-03	12	4.51E-01	0.954	3.90E-03	<i>PANK2*</i>
10	22842590	rs943189	A	6.12E-01	1.022	6.86E-01	12	8.82E-02	0.94	3.59E-02	12	8.82E-02	0.94	3.59E-02	<i>SPAG6 and LOC643475</i>
12	73284781	rs4760851	A	6.15E-01	1.021	9.57E-01	12	7.51E-02	0.94	6.50E-02	12	7.51E-02	0.94	6.50E-02	<i>TRHDE and LOC100128674</i>
3	1940922	rs1387089	G	6.80E-01	0.972	4.43E-01	11	1.13E-02	0.861	3.64E-02	11	1.13E-02	0.861	3.64E-02	<i>LOC391504 and LOC727810</i>
22	47687170	rs1005239	G	7.07E-01	0.984	5.10E-01	12	1.44E-02	0.918	6.50E-02	12	1.44E-02	0.918	6.50E-02	<i>TBC1D22A and RP11-191L9.1</i>
21	27774870	rs2032257	A	7.14E-01	1.015	3.24E-01	12	7.68E-02	0.94	9.90E-03	12	7.68E-02	0.94	9.90E-03	<i>APP and CYPR1</i>
8	115504434	rs7836298	G	7.16E-01	1.026	6.19E-01	12	5.60E-02	0.889	7.20E-03	12	5.60E-02	0.889	7.20E-03	<i>hCG_1644355 and TRPS1</i>
12	73441110	rs952394	G	7.31E-01	0.986	9.15E-01	12	5.54E-02	1.068	8.47E-02	12	5.54E-02	1.068	8.47E-02	<i>TRHDE and LOC100128674</i>
12	12010736	rs2255953	G	8.05E-01	1.013	8.06E-01	11	5.12E-03	1.131	4.54E-02	11	5.12E-03	1.131	4.54E-02	<i>ETV6*</i>
16	19904082	rs13333521	A	8.34E-01	1.022	6.71E-02	9	3.11E-01	1.181	2.70E-03	9	3.11E-01	1.181	2.70E-03	<i>GPRC5B and GPR139</i>
22	30088257	rs1009147	A	8.58E-01	0.989	9.91E-01	11	2.07E-02	0.881	2.06E-01	11	2.07E-02	0.881	2.06E-01	<i>NFZ*</i>
8	144974963	rs7831595	A	8.96E-01	1.005	1.82E-01	12	1.90E-02	1.084	5.90E-03	12	1.90E-02	1.084	5.90E-03	<i>EPPK1</i>
11	22379184	rs11026531	A	9.75E-01	1.002	3.72E-01	11	3.49E-02	0.916	2.31E-02	11	3.49E-02	0.916	2.31E-02	<i>SLC17A6*</i>

P-value is fixed p-value if p-value for heterogeneity (Q) > 0.005, and random p-value if p-value for heterogeneity (Q) < 0.005, and random OR if p-value for heterogeneity (Q) > 0.005, and random OR if p-value for heterogeneity (Q) < 0.005; OR is Odds Ratio; OR is fixed OR if p-value for heterogeneity (Q) > 0.005, and random OR if p-value for heterogeneity (Q) < 0.005; Q is p-value for heterogeneity; N = number of cohorts;

\*means that the corresponding SNP is an intron in this gene.

doi:10.1371/journal.pone.0091621.t006



**Figure 3. Forest plot showing evidence of association for rs6577641 with chronic mucus hypersecretion in the identification and replication cohorts.** Vertically left, the identification cohort and the replication cohorts included in the meta-analysis. The boxes represent the precision and the horizontal lines represent the confidence intervals. The squares represent the pooled effect estimate from the meta-analysis of all cohorts. The horizontal axis shows the scale of the effects.

doi:10.1371/journal.pone.0091621.g003

17% per G allele to develop CMH in a population of ex- and current heavy smokers.

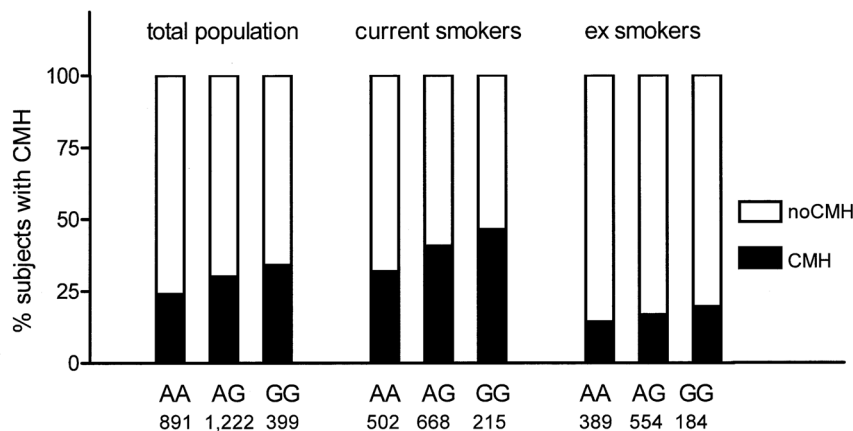
Multivariate regression analysis, stratified for current an ex-smoking, showed essentially the same effect sizes and direction of the association of CMH and the risk allele of rs6577641. It is likely

that lack of power is the reason for not reaching the level of significance in ex-smokers.

These data strongly suggest that *SATB1* plays a role in the susceptibility to CMH in subjects with a history of heavy smoking ( $\geq 20$  pack-years) within the general population. Moreover, rs6577641 has a *cis*-eQTL effect on *SATB1* lung tissue expression, the risk allele at rs6577641 (G) increasing and the A-allele reducing expression of *SATB1* significantly. Additionally, we found a higher *SATB1* expression in bronchial biopsies of COPD-patients with CMH. We found no differences between the GG and AA genotypes for protein expression of *SATB1* in airway epithelium by IHC in a small sample from our lung tissue registry. Finally, we demonstrate that *SATB1* mRNA expression is induced during mucociliary differentiation in ALI cultures of human bronchial epithelial cells of 2 donors supporting our eQTL findings. Interestingly, expression of the mucin gene *MUC5AC* was also induced during this culture period, with a slightly delayed kinetics compared to *SATB1*. Together these data strongly suggest that *SATB1* is induced during differentiation of bronchial epithelial cells and affects chronic mucus hypersecretion.

The forest plot clearly shows that the effect of SNP rs6577641 is lower in cohorts including COPD patients only (GLUCOLD, Rucphen, COPDGene, ECLIPSE and Norway) than in the other cohorts. Additional meta-analysis of COPD-cohorts and general population based cohorts separately confirmed this (COPD cohorts, combined p-value = 0.236, OR = 1.07 and general population based cohorts, combined p-value =  $5.18 \times 10^{-7}$ , OR = 1.26). This suggests genetic heterogeneity of CMH in subjects with and without COPD.

The SNP most significantly associated with CMH, rs6577641, is located in an intron of *SATB1*. *SATB1* is a transcription factor and chromatin (re)organizer important for controlling the expression of many genes in a tissue or cell-type specific fashion, for instance in differentiating thymus T-cells [17] or differentiating skin keratinocytes [18]. Expression of *SATB1* has been observed in normal human bronchial epithelial cells by immunohistochemistry and lower levels were observed in non-small lung cancer cells [16]. In our study, we also showed the presence of *SATB1* in bronchial epithelial cells by IHC staining of lung tissue. However, no significant differences were found between patients homozygous for the protective and risk alleles, for either specific *SATB1* staining or for PAS staining, the latter specifically detecting mucus. This inability to detect a genotype effect on protein staining may



**Figure 4. Percentage of subjects with chronic mucus hypersecretion (CMH) within genotypes (AA, AG and GG) of rs6577641 in the identification cohort (NELSON), and distributed among ex- and current smokers.**

doi:10.1371/journal.pone.0091621.g004

**Table 7.** Meta-analysis of the effect of rs6577641 on mRNA expression levels of *SATB1* in the lung\*.

Probe Gene Symbol	Affymetrix Probe ID	Z-score Groningen	Z-score Laval	Z-score UBC	Z-Score Meta-Analysis	p-value Meta-Analysis
		N = 351	N = 335	N = 409		
<i>SATB1</i>	100148784_TG1_at	-2.28	-0.08	-1.62	-2.29	<b>0.022</b>
<i>SATB1</i>	100150253_TG1_at	-0.84	-0.49	-1.62	-1.70	0.088
<i>SATB1</i>	100305926_TG1_at	-2.81	-1.38	-1.46	-3.26	<b>0.001</b>

\*To assess the effect of the SNP rs6577641 on gene expression, a Kruskal-Wallis test was performed. This test generates a p-value, but does not give a direction of the effect. To assess the direction of the effect, a Spearman's correlation test was performed. Next, a Z-score was calculated for each center and a meta-analysis performed for each of the three *SATB1* probes across all centers. Finally, a meta-analysis for all three *SATB1* probes was performed across all centers. This generated a Z-score of  $-5.87$  and a corresponding p-value of  $4.3 \times 10^{-9}$ , indicating that the susceptibility G allele of the SNP rs6577641 increases *SATB1* expression. doi:10.1371/journal.pone.0091621.t007

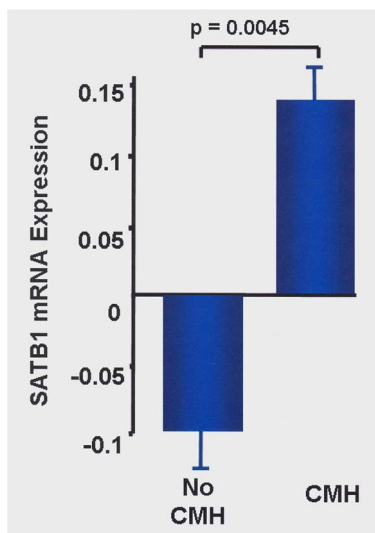
be due to lack of power, as we found a large variation in *SATB1* and PAS protein expression in the relatively small number of lung tissue samples. Other explanations include possible expression regulation of *SATB1* by smoke exposure which could be a dynamic process not readily detected at the protein level by any single-time point analysis such as IHC staining on lung biopsies. Alternatively *SATB1* expression levels may vary throughout the lungs or the technique used here is not sensitive enough to detect relatively small differences in protein levels. To further explore the association of *SATB1* protein and its underlying regulation, it would be of interest to perform longitudinal investigations on lung tissue samples of subjects with and without CMH, or time series of *in vitro* cultured epithelial cells from donors with a specific genotype and cigarette smoke exposure. This would also allow further studies on epigenetic regulation with methylation, microRNA or histone modifications.

The lack of association between the *SATB1* protein and rs6577641 might additionally be due to the location of mucus positive cells in lung tissue. Mucus is produced both by goblet cells and submucosal glands, which we did not investigate further. Normal mucus consists of 97% water and 3% solids including 30% mucins. In case of dysregulation of mucus production, the

concentration of solids in mucus may increase up to 15%. A further step therefore could involve investigating mucins/proteins present in mucus, e.g. MUC5AC is predominantly produced by goblet cells in proximal airways and MUC5B by secretory cells throughout the airways and by submucosal glands.

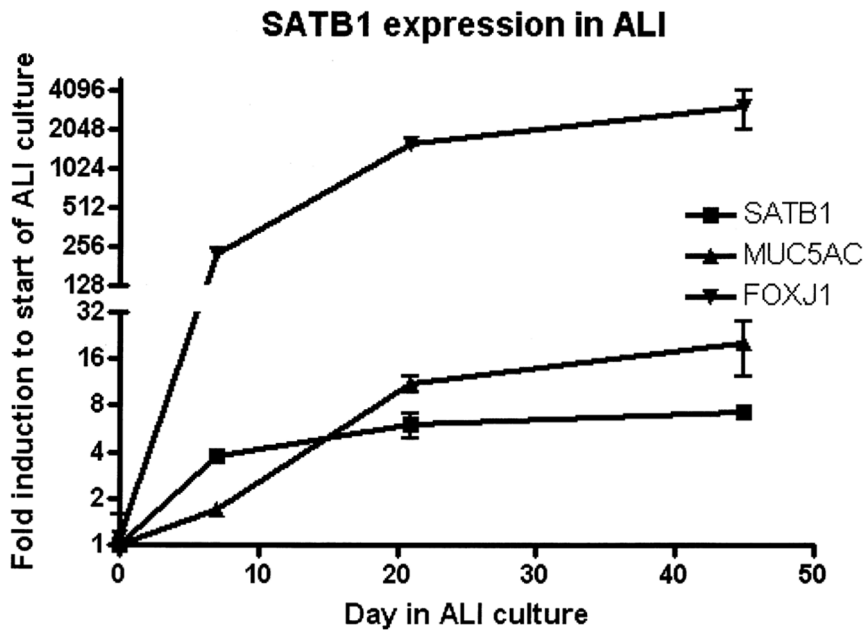
How does *SATB1* expression contribute to CMH? *SATB1* is known to be a genome organizer, a tissue specific chromatin remodeling protein with a property to modifying chromatin architecture by formation of loops, allowing contact of condensed genomic DNA to regulatory transcription proteins [19]. Thus *SATB1* can control gene expression of a series of target genes located within a single locus at a specific chromosomal location [20]. This has for instance been elegantly shown in case of differentiating keratinocytes [18], where *Satb1* expression regulates genes located in the keratinocyte-specific loci, leading to adaptation of a specific cell fate of the differentiating keratinocytes. Similarly, a mechanism by which *SATB1* could contribute to CMH is the induction of a gene expression program during differentiation of bronchial epithelial cells, leading to adaptation of a cell fate specific for mucus producing cells in the submucosal glands or a goblet cell phenotype in the bronchial epithelium. Involvement of *Satb1* in pneumocyte differentiation was previously observed by Baguma et al. in mice [21]. We observed induction of *SATB1* expression in bronchial epithelial cells differentiating under ALI culture conditions. Further research will need to test whether a specific gene expression profile is induced by *SATB1* expression in differentiating bronchial epithelial cells. *SATB1* is also highly expressed in thymocytes, but absent in mature non-activated T cells [22]. Moreover, *Satb1* has been shown in mice to be essential for expression of  $T_{\text{helper}2}$  ( $Th2$ ) cells important in the regulation of genes encoding interleukin 4, 5 and 13 [19]. In *Satb1*-deficient mice, development of thymocytes stopped after the  $CD4^+/CD8^+$  stage with deregulation of many genes [23]. Conversely, in case of excessive *SATB1*-production an excess of  $Th2$  cells may be formed which all produce IL-13, which may contribute to increased mucus production. Therefore, a putative role of *SATB1* in T-cells for the CMH phenotype should not be disregarded.

Strength of our study is the fact that we were able to replicate our findings in different populations, ranging from cohorts consisting of individuals with severe airflow limitation to cohorts mainly consisting of healthy smokers. There are some limitations, e.g. the presence of CMH was not based on actual measurements of the amount of sputum produced but based on questionnaires that were not completely similar in all study cohorts. Underreporting of CMH occurs since those experiencing CMH become accustomed to these symptoms, believing they are smoking related or because they are embarrassed to admit to cough and sputum.



**Figure 5.** Bronchial biopsy mRNA-expression levels of *SATB1* in COPD patients with chronic mucus hypersecretion (n = 38) compared to patients without chronic mucus hypersecretion (n = 39).

doi:10.1371/journal.pone.0091621.g005



**Figure 6.** *SATB1*, *MUC5AC* and *FOXJ1* mRNA expression levels during mucociliary human airway epithelial cell differentiation (n = 2 donors). Expression of *SATB1*, the identified gene in our study, *MUC5AC* a marker of mucus, and *FOXJ1*, representing ciliated cells in epithelial cell culture on air liquid interface.

doi:10.1371/journal.pone.0091621.g006

We demonstrated that *SATB1* mRNA expression is induced during mucociliary differentiation in ALI cultures of HBE cells in a small dataset (n = 2). However, these data seem reliable as they are supported by eQTL data from lung tissue. Despite this drawback, we consistently found evidence for association of *SATB1* with CMH in the populations studied, showing the robustness of our finding. Moreover, we corroborated this finding by functional studies in lung tissue, airway wall biopsies of COPD patients and epithelial cultures. More extensive research is needed to investigate which factors induce *SATB1* expression in airway epithelium.

In summary, we performed identification analyses and meta-analyses using data from almost 7,000 participants to identify genes involved in susceptibility for CMH. It is remarkable that we found a genetic association for CMH given this phenotype is partly subjectively determined and not well delineated. Moreover, despite cohort differences to define CMH and severity of airflow limitation, we found consistent effects of SNP rs6577641 on CMH. This confirms that the CMH phenotype, despite the fact that it is self-reported, is a robust phenotype irrespective of the presence or absence of airflow limitation. The association of rs6577641 on chromosome 3 at the *SATB1* locus with CMH was supported by functional studies including gene expression findings, demonstrating *SATB1* to be associated with CMH.

Chronic mucus hypersecretion is a bothersome symptom for many people, it increases in prevalence with aging and affects quality of life, exacerbations of symptoms due to respiratory infections and ultimately increases mortality. The involvement of *SATB1* in CMH offers opportunities to better understand the process leading to CMH, and future development of tailored medicines.

## Supporting Information

**Supplement S1**  
(DOC)

## Acknowledgments

The authors would like to thank the research staff at Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance in the lung eQTL study.

LifeLines Cohort Study: BZ Alizadeh<sup>1</sup>, RA de Boer<sup>2</sup>, HM Boezen<sup>1</sup>, M Bruinenberg<sup>3</sup>, L Franke<sup>4</sup>, P van der Harst<sup>5</sup>, HL Hillege<sup>1,2</sup>, MM van der Klauw<sup>5</sup>, G Navis<sup>6</sup>, J Ormel<sup>7</sup>, DS Postma<sup>8</sup>, JGM Rosmalen<sup>7</sup>, JP Slaets<sup>9</sup>, H Snieder<sup>1</sup>, RP Stolk<sup>1</sup>, BHR Wolffenbuttel<sup>5</sup>, C Wijmenga<sup>4</sup>

<sup>1</sup>University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands;

<sup>2</sup>University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, the Netherlands;

<sup>3</sup>University of Groningen, University Medical Center Groningen, the LifeLines Cohort Study, Groningen, the Netherlands;

<sup>4</sup>University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands;

<sup>5</sup>University of Groningen, University Medical Center Groningen, Department of Endocrinology, Groningen, the Netherlands;

<sup>6</sup>University of Groningen, University Medical Center Groningen, Department of Internal Medicine, Division of Nephrology, Groningen, the Netherlands;

<sup>7</sup>University of Groningen, University Medical Center Groningen, Interdisciplinary Center of Psychopathology of Emotion Regulation (ICPE), Department of Psychiatry, Groningen, the Netherlands;

<sup>8</sup>University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands

## Author Contributions

Conceived and designed the experiments: AED DSP HMB CW JS PZ HG JWJ WT MCN MvdB. Performed the experiments: AED JS ML MCN UB MP. Analyzed the data: AED MvdB DL WT MHC MCN MP ML MB FM M. Obeidat. Contributed reagents/materials/analysis tools: ML MP JS M. Obeidat MvdB DL. Wrote the paper: AED DSP HMB CW JMV GHK PJS EKS MHC MCN. Acquisition of replication cohorts: M. Dahlback KT PSH PJS AS JV BGN M. Dahl WMV HSP HAS M. Owsjiewitsch HUK HJK EN PN CCvD EKS JDC THB DAL PB AG YB LL FR AGU AH BHS GGB CMvD UB GHK SN.

## References

- Anonymous (1995) Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. american thoracic society. *Am J Respir Crit Care Med* 152: S77–121.
- Ferre A, Fuhrman C, Zureik M, Chouaid C, Vergnenegre A, et al. (2012) Chronic bronchitis in the general population: Influence of age, gender and socio-economic conditions. *Respir Med* 106: 467–471.
- Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, et al. (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 176: 532–555.
- Vestbo J, Rasmussen FV (1989) Respiratory symptoms and FEV1 as predictors of hospitalization and medication in the following 12 years due to respiratory disease. *Eur Respir J* 2: 710–715.
- Vestbo J, Prescott E, Lange P (1996) Association of chronic mucus hypersecretion with FEV1 decline and chronic obstructive pulmonary disease morbidity. copenhagen city heart study group. *Am J Respir Crit Care Med* 153: 1530–1535.
- World Health Organization website. Available: [Http://www.who.int/respiratory/copd/burden/en/index.html](http://www.who.int/respiratory/copd/burden/en/index.html). Accessed January 2013.
- Shurin S (2012) Morbidity & mortality: 2012 chart book on cardiovascular, lung, and blood diseases. 2013: 107.
- Lange P, Parner J, Prescott E, Vestbo J (2003) Chronic bronchitis in an elderly population. *Age Ageing* 32: 636–642.
- Ekberg-Aronsson M, Pehrsson K, Nilsson JA, Nilsson PM, Lofdahl CG (2005) Mortality in GOLD stages of COPD and its dependence on symptoms of chronic bronchitis. *Respir Res* 6: 98.
- Viegi G, Carrozzi L, Di Pede F, Baldacci S, Pedreschi M, et al. (1994) Risk factors for chronic obstructive pulmonary disease in a north italian rural area. *Eur J Epidemiol* 10: 725–731.
- Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, et al. (1998) Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. risk to relatives for airflow obstruction and chronic bronchitis. *Am J Respir Crit Care Med* 157: 1770–1778.
- Hallberg J, Dominicus A, Eriksson UK, Gerhardsson de Verdier M, Pedersen NL, et al. (2008) Interaction between smoking and genetic factors in the development of chronic bronchitis. *Am J Respir Crit Care Med* 177: 486–490.
- Zhu G, Agusti A, Gulsvik A, Bakke P, Coxson H, et al. (2009) CTLA4 gene polymorphisms are associated with chronic bronchitis. *Eur Respir J*.
- Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, et al. (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 8: e1003029.
- Lapperre TS, Snoeck-Stroband JB, Gosman MM, Jansen DF, van Schadewijk A, et al. (2009) Effect of fluticasone with and without salmeterol on pulmonary outcomes in chronic obstructive pulmonary disease: A randomized trial. *Ann Intern Med* 151: 517–527.
- Selinger CI, Cooper WA, Al-Sohaily S, Mladenova DN, Pangon L, et al. (2011) Loss of special AT-rich binding protein 1 expression is a marker of poor survival in lung cancer. *J Thorac Oncol* 6: 1179–1189.
- Cai S, Lee CC, Kohwi-Shigematsu T (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* 38: 1278–1288.
- Fessing MY, Mardaryev AN, Gdula MR, Sharov AA, Sharova TY, et al. (2011) p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194: 825–839.
- Cai C, Zhang HY, Le JJ, Dong JC, Cui Y, et al. (2010) Inflammatory airway features and hypothalamic-pituitary-adrenal axis function in asthmatic rats combined with chronic obstructive pulmonary disease. *Chin Med J (Engl)* 123: 1720–1726.
- Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419: 641–645.
- Baguma-Nibasheka M, Angka HE, Inanlou MR, Kablar B (2007) Microarray analysis of *Myf5*<sup>-/-</sup>:*MyoD*<sup>-/-</sup> hypoplastic mouse lungs reveals a profile of genes involved in pneumocyte differentiation. *Histol Histopathol* 22: 483–495.
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T (1992) A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* 70: 631–645.
- Krangel MS (2007) T cell development: Better living through chromatin. *Nat Immunol* 8: 687–694.
- van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, et al. (2009) Management of lung nodules detected by volume CT scanning. *N Engl J Med* 361: 2221–2229.
- 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.
- 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.
- 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.
- Rijcken B, Schouten JP, Mensinga TT, Weiss ST, De Vries K, et al. (1993) Factors associated with bronchial responsiveness to histamine in a population sample of adults. *Am Rev Respir Dis* 147: 1447–1453.
- van Diemen CC, Postma DS, Vonk JM, Bruinenberg M, Nolte IM, et al. (2006) Decorin and TGF-beta1 polymorphisms and development of COPD in a general population. *Respir Res* 7: 89.
- Verschuren WM, Blokstra A, Picavet HS, Smit HA (2008) Cohort profile: The doetinchem cohort study. *Int J Epidemiol* 37: 1236–1241.
- Nizankowska-Mogilnicka E, Mejza F, Buist AS, Vollmer WM, Skucha W, et al. (2007) Prevalence of COPD and tobacco smoking in malopolska region—results from the BOLD study in poland. *Pol Arch Med Wewn* 117: 402–410.
- Lamprecht B, McBurnie MA, Vollmer WM, Gudmundsson G, Welte T, et al. (2011) COPD in never smokers: Results from the population-based burden of obstructive lung disease study. *Chest* 139: 752–763.
- Becker N, Delorme S, Kauczor HU (2008) LUSI: The german component of the european trial on the efficacy of multislice-CT for the early detection of lung cancer. *onkologie* 31.
- Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, et al. (2008) Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J* 31: 869–873.
- Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, et al. (2010) Genetic epidemiology of COPD (COPDGene) study design. *COPD* 7: 32–43.
- Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, et al. (2011) A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*.
- Grydeland TB, Dirksen A, Coxson HO, Pillai SG, Sharma S, et al. (2009) Quantitative computed tomography: Emphysema and airway wall thickness by sex, age and smoking. *Eur Respir J* 34: 858–865.