Lung Function in the General Population: the Complex Interplay of Variants in *SERPINA1* and other Genes with the Environment

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Gian Andri Thun

aus Ramosch (GR)

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Marcel Tanner, Prof. Dr. Nicole Probst-Hensch, Prof. Dr. Stefano
Guerra

Basel, den 15. Oktober 2013

Prof. Dr. Jörg Schibler

Dekan der Philosophisch-
Naturwissenschaftlichen Fakultät

Table of Contents

Zusammenfassung

Hintergrund. Unter den weltweit häufigsten Gesundheitsproblemen befinden sich obstruktive respiratorische Erkrankungen wie Asthma oder chronisch obstruktive Lungenerkrankung (COPD). Beide zeigen sehr heterogene Erscheinungsformen und werden meistens durch Lungenfunktionsmessungen diagnostiziert. Neben etablierten Umweltrisiken gibt es auch genetische Faktoren, welche die Lungenfunktion wesentlich mitbestimmen. Insbesondere *SERPINA1* Genvarianten, welche die Konzentration von Alpha1-Antitrypsin (AAT) im Blut stark reduzieren und folglich zu einem gestörten Gleichgewicht von Proteasen und Antiproteasen in der Lunge führen, sind seit vielen Jahren als Risikofaktor bekannt. Hingegen sind Genvarianten, welche zu einem schwachen oder intermediär ausgeprägten Mangel an AAT im Blut führen, vermutlich nur für einen Teil der Bevölkerung ein Risikofaktor, aber weder wissen wir genau, wie der Konzentrationsbereich eines solchen Mangels am besten zu definieren ist, noch welche Zusatzfaktoren gesundheitsrelevant für betroffene Personen sind.

Methoden. In dieser Arbeit wurde mithilfe der schweizerischen Kohortenstudie über Luftverschmutzung und Lungenerkrankungen bei Erwachsenen (SAPALDIA) ermittelt, welche genetischen Polymorphismen die AAT-Serumkonzentration hauptsächlich bestimmen. Im Weiteren wurden die Konzentrationsbereiche einer AAT-Defizienz für die Allgemeinbevölkerung neu festgelegt und die Assoziation zwischen *SERPINA1* Defizienzgenotypen und altersbedingter Lungenfunktionsabnahme umfassend untersucht. Die Beurteilung von Gen-Umwelt-Interaktionen in Bezug auf respiratorische Gesundheit war dabei ein zentraler Teil der Arbeit und umfasste neben *SERPINA1* auch weitere Gene. Der umweltbezogene Fokus wurde auf Rauchen, Luftverschmutzung, berufliche Exposition gegenüber Dämpfen, Gas, Staub und Rauch sowie Fettleibigkeit gelegt; alles Faktoren, die in Verdacht stehen, Entzündungen zu begünstigen. Da in SAPALDIA Blutwerte von Entzündungsmarkern inklusive AAT, genomweite Daten inklusive zusätzlicher Genotyp- und Sequenzdaten des *SERPINA1* Gens sowie umfassende und detaillierte Umwelt- und respiratorische Gesundheitsdaten erhoben wurden, war die Studie wie keine andere prädestiniert, die beschriebenen Zusammenhänge und Interaktionen zu untersuchen.

Resultate. Wir beobachteten eine homogenere Verteilung von AAT-Serumkonzentrationen in Personen mit intermediär ausgeprägter AAT-Defizienz als aus der Literatur bekannt und

verdeutlichten die Abhängigkeit der AAT-Blutwerte von anderen Entzündungsmarkern. Weiter konnten wir bestätigen, dass seltene Varianten des *SERPINA1* Gens die hauptsächlichen genetischen Determinanten der AAT-Blutwerte sind und wiesen auf einige der inhärenten Schwächen von genomweiten Assoziationsstudien hin. Eine hohe Belastung mit entzündungsfördernden Einflüssen modifizierte die Assoziation zwischen intermediär ausgeprägter AAT-Defizienz und Lungenfunktionsabnahme. Genetische Interaktionen mit Fettleibigkeit hinsichtlich Asthma und mit Luftverschmutzung hinsichtlich Lungenfunktionsabnahme wurden ebenfalls gefunden, was auf der einen Seite das immer noch reichhaltige Forschungsgebiet der Gen-Umweltinteraktionen aufzeigt, welches noch nicht systematisch untersucht worden ist, aber auf der anderen Seite die Komplexität offenbart, allgemeingültige Schlussfolgerungen aus solchen Analysen ziehen zu können.

Diskussion und Schlussfolgerung. Zusammenfassend lässt sich sagen, dass diese Arbeit möglicherweise den Ablauf des diagnostischen Verfahrens vereinfacht, mit dem sich Patienten mit einer vermuteten AAT-Defizienz konfrontiert sehen. Obwohl Personen mit einer intermediär ausgeprägten AAT-Defizienz im Allgemeinen nicht als eine Risikogruppe für respiratorische Gesundheitsbeeinträchtigungen angeschaut werden, scheinen sie anfälliger als die Allgemeinbevölkerung auf entzündliche Stressoren zu sein. So würden sie möglicherweise stärker von Massnahmen wie Beratungen gegen die Aufnahme des Rauchens, für gesunde Ernährung oder für verbesserte betriebliche Sicherheit.

Summary

Background. One of the globally most frequent health problems are obstructive lung diseases such as asthma and chronic obstructive pulmonary disease. Both of them show heterogeneous phenotypes and are most commonly diagnosed by lung function measurements. Apart from several well-established environmental risk factors, there are also genetic factors which play an important role in determining lung function. Notably, *SERPINA1* gene variants which severely reduce the alpha1-antitrypsin (AAT) concentration in the blood and consequently lead to a protease-antiprotease disequilibrium in the lung have been known as risk factors for several years. Intermediate deficiency of AAT serum level is however assumed to be a risk factor in only part of the population, but neither is it entirely clear how to define this range of protein concentration for the general population, nor do we know which co-factors are health-relevant in intermediately deficient individuals.

Methods. In this work, SAPALDIA, the Swiss cohort study on air pollution and lung disease in adults, was used to find the essential genetic polymorphisms which determine AAT serum level. Deficiency ranges for AAT were defined in the general population, and the association between *SERPINA1* deficiency genotypes and age-related lung function decline was investigated in a comprehensive way. The assessment of gene-environment interactions in terms of pulmonary health was a central part of this work and embraced also genes beyond *SERPINA1*. The environment-related focus was set on factors associated with inflammatory stress, namely smoking, air pollution, high occupational exposure to vapours, gas, dusts and fumes as well as obesity. The availability of serum inflammatory markers including AAT, genome-wide data including additional genotype and sequence information of the *SERPINA1* gene as well as very comprehensive and detailed environmental and respiratory health data made SAPALDIA, unlike any other study, ideally suited to investigate the aforementioned associations and interactions.

Results. This work found a smaller range of AAT serum level in subjects with intermediate AAT deficiency than reported in the literature and clarified the role of elevated inflammatory conditions on AAT serum level. It confirmed uncommon variants in the *SERPINA1* locus as the major genetic determinants of AAT blood level and pointed to some of the inherent weaknesses of genome-wide association studies. A high burden of inflammatory stress was suggested to modify the association between intermediate AAT

deficiency and lung function decline. Further genetic interaction with obesity in terms of asthma and with air pollution in terms of lung function decline was suggested, pointing on the one hand to a still proliferative research area of gene-environment interactions which has not yet been systematically assessed, but revealing on the other hand the complexity of drawing firm conclusions from such analyses.

Discussion and Conclusion. In summary, this work may potentially facilitate the diagnostic procedure for subjects with an assumed AAT deficiency. Although generally not regarded as a risk group for adverse pulmonary health, individuals with an intermediate AAT deficiency seem more susceptible to elevated inflammatory conditions compared to the general population. They would potentially more strongly benefit from measures like counselling against the uptake of smoking, for healthy diet programmes or improvements of occupational safety.

Abbreviations

AAT	Alpha1-Antitrypsin
AUC	Area Under the Curve
BMI	Body Mass Index
COPD	Chronic Obstructive Pulmonary Disease
CVD	Cardio-Vascular Disease
CNV	Copy Number Variation
CRP	C-Reactive Protein
DALY	Disability-Adjusted Life Year
DTC	Direct To Consumer
eQTL	Expression Quantitative Trait Locus
ETS	Environmental Tobacco Smoke
$FEF_{25-75\%}$	Forced Expiratory Flow between 25 and 75% of FVC
FEV_1	Forced Expiratory Volume in one Second
FVC	Forced Vital Capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GWAS	Genome-Wide Association Study
GWIS	Genome-Wide Interaction Study
LD	Linkage Disequilibrium
LMIC	Low- and Middle-Income Countries
MAF	Minor Allele Frequency
NCD	Non-Communicable Disease
PCR	Polymerase Chain Reaction
PM_{10}	Particulate Matter < 10 µm
RCT	Randomised Controlled Trial
SAPALDIA	Swiss Cohort Study on Air Pollution and Lung Disease in Adults
SNP	Single Nucleotide Polymorphism
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Acknowledgements

My deepest gratitude goes to my supervisor Nicole Probst-Hensch for accepting me as her PhD student, for her contagious enthusiasm in genetic and environmental research, for continuous support during the whole PhD and for finding a very fair balance of motivation, promotion and demand.

I am very thankful to Medea Imboden who was my pre-PhD supervisor for introducing me to the organisation of the SAPALDIA biobank, the genotyping and other wet lab work and to first data analyses in epidemiology. She contributed greatly to my decision of extending the laboratory position to an employment as a PhD student.

Sincere thanks go to Ivan Curjuric and Martin Adam for being great collaborators and for numerous stimulating discussions and constant support.

My practical work was carried out in Pavia for which I want to thank Maurizio Luisetti, Ilaria Ferrarotti and the team of the alpha1-antitrypsin registry for their cooperative and supportive attitude.

I am grateful to Stefano Guerra from CREAL, Marcel Tanner, the Swiss TPH director, and Maurizio Luisetti for serving as the co-referee, the faculty representative and the main expert in my PhD committee.

I acknowledge all participants for their willingness to contribute to SAPALDIA and all study members involved in designing and conducting this prolific study, which was the prerequisite for the high quality data I could use for my analyses. Financial support came from the Swiss National Foundation and Talecris GmbH, which made this PhD possible.

Last but not least, I want to thank my parents for their unconditional support and encouragement during all stages of my education.

1 Background

1.1 Setting the Stage: Non-Communicable Disease Research in the Framework of Genetic Epidemiology

Non-communicable diseases (NCDs) are medical conditions which are not directly transmissible from person to person and usually of slow progression, long duration and therefore age-related. The term "chronic disease" is often used interchangeably, but this is not entirely correct as some chronic diseases may also be caused by infections. Major types of NCDs include cardiovascular diseases (CVDs), cancers, respiratory diseases, auto-immune diseases as well as endocrinal and neurological disorders.

Genetic epidemiology is a research field which concentrates on studying the role of genetic variation and its interaction with the environment in determining health and disease at the population level. Comparably novel instruments have recently revolutionised this discipline and will therefore be presented in sufficient depth in the first part of the introductory chapter. The second part is then dedicated more specifically to the traits and diseases which were actually investigated during this PhD work.

1.1.1 Global Impact and Aetiology of Non-Communicable Diseases

The global burden of disease estimates show that the burden of NCDs has steadily increased over the past 20 years. In 2010, NCDs accounted for more than 50% of all disability-adjusted life years (DALYs) [1], a measure that adds the years of life lost to those lived with disability. Most prominently ranked are the diseases of blood circulation, cancers and chronic respiratory diseases. The proportion of NCDs on global DALYs is expected to further rise in future, in particular owing to the increasing life expectancy of the world's population, the increasing rates of risk factors like smoking and overweight and the better management of communicable diseases.

The aetiology of NCDs is usually complex and often described as depending on a heritable genetic component as well as on the non-heritable environment. However, this distinction is not very clear-cut. Genetics also consists of a non-heritable part (e.g. the postzygotic mutations crucial in the aetiology of many cancers), and the environment may lead under

certain circumstances to heritable modifications in germ cells (e.g. by inducing epigenetic changes during development). Nevertheless, when making the gene vs. environment distinction, we restrict the genetic component to its heritable part and regard the environment as a proxy for all non-heritable influences including also stochastic events.

1.1.2 Environment

Important factors include behavioural and social determinants beside factors of the in- and outdoor ambience. Globally, it was estimated that the highest-ranking risk factors based on DALYs for 2010 were high blood pressure, tobacco smoking and household air pollution [2]. A large part of the environmental component is modifiable and therefore of high public health relevance. In practice, however, it has turned out to be difficult to modify the distribution of some of the most relevant environmental risk factors (e.g. tobacco smoking).

It is important to note that the inference of causality in any given association between an environmental factor and a disease is all but a trivial endeavour. Observational studies typically investigating the impact of such factors (exposures) on certain traits or diseases (outcomes) are renowned for the high risk of detecting spurious associations. Different types of biases, confounding and reverse causality issues have to be taken into account. This can be best countered by using a randomised controlled trial (RCT) study design, but this is often not practically or ethically feasible. Cohort studies can at least overcome some of these pitfalls due to the temporal sequence of cause and effect. Other important challenges are the generally long latency period between cause and effect and the often inaccurate or unreliable measurement (or allocation) of environmental factors, additionally complicated by time-dependency. Susceptibility to a certain exposure may depend on unknown time windows or on the duration of exposure [3]. One approach sometimes used to infer causality is Mendelian randomisation. If a genetic factor can serve as a proxy for an environmental exposure, and if the genetic factor is additionally associated with the respective outcome, we could infer causality to the association between the exposure and the outcome. This is because genetic factors are not believed to be associated with possible factors that confound the association. Nevertheless, the method is only useful if the environmental factor is a biomarker or a behaviour (for which we normally find genetic surrogates) and if pleiotropy (i.e. if the genetic factor is linked via different pathways to the exposure and to the outcome) is unlikely [4].

1.1.3 Heritability

The proportion of a trait's phenotypic variation which is due to heritable genetic factors is defined as the trait's (broad-sense) heritability. This term is normally defined for a specific population at a specific age and in a particular environment [5]. Non-additive components of the heritable factors are interaction effects either between alleles at the same locus (dominance) or at different loci (gene-gene interactions, epistasis). They are difficult to assess, but are believed to be minor compared to the additive components, i.e. the additive effects of the alleles [6,7]. Therefore, one usually refers to the narrow-sense heritability, which is the ratio between the additive components of the heritable factors and the total phenotypic variation. Several methods are in use to estimate this proportion including quantification of trait or disease correlation in pairs of relatives of a specific type or assessment of differences among monozygotic and dizygotic twins [8]. Such family- or pedigree-based methods have the advantage that they do not have to consider population heterogeneity, but they are usually not very powerful and struggle with shared environment issues. Another method to get a measure of the additive genetic variance uses the effects of all genetic variants in total, as typically provided by a genome-wide association study (GWAS, see 1.1.5) [9]. Such a measure is likely an underestimation of the true value and can serve as a lower limit of the heritability since independent effects of rare variants are not taken into account.

Heritability estimates from pedigree studies typically lie in the range of 0.3 to 0.9 (30 – 90%) for many common metabolic and anthropomorphic traits as well as for a range of metabolic, auto-immune and neurological diseases [10,11]. This apparently contrasts with the high importance of environmental risk factors in the aetiology of many NCDs, which is for instance evident by the steep increase of such diseases in certain migrant populations [12]. Heritability estimates depend however on the distribution of environmental factors and should therefore ideally be reported according to sex, age, area and population. Additional caution in the interpretation of heritability estimates is required due to the different estimation methods.

1.1.4 Identification of Genes: The Candidate Gene Approach

Genetic linkage studies, in which the chromosomal segregation is compared with the disease segregation in families, proved unsuitable to detect genetic variants associated with complex

diseases. Hence, the approach which was normally used until the year 2006 was the investigation of associations between variants of candidate genes and a specific outcome. Compared to environmental risk factors (see 1.1.2), genetic risk factors are in many ways easier to deal with in association studies. They inherently antedate any disease outcome, they do not change during lifetime, and they can be measured with high accuracy and reliability. Since genetic polymorphisms assort randomly during meiosis, environmental factors (with the exception of population heterogeneity) are not expected to confound genetic associations with traits or diseases. Adjusting for environmental factors is therefore hardly crucial, but makes sense in the case of behavioural factors since they may lie on the causal pathway between the genetic risk factor and the disease.

Nevertheless, replication studies and meta-analyses showed that reported associations of genetic variants, usually single nucleotide polymorphisms (SNPs), could rarely be confirmed [13]. This had likely to do with the fact that the investigated SNPs were not necessarily the causal ones, but correlated with them (i.e. lying in linkage disequilibrium, LD), and the correlation structure differs across populations. There were also other reasons for non-replication. Most importantly, confounding by population heterogeneity (because allele as well as disease frequencies deviate considerably across populations) might often have led to spurious association within studies. In addition, publication bias was assumed to be common, and since the number of potentially functional genetic variants per gene is large, selective reporting might have occurred. Studies were often small with little power to detect SNPs with small effects, and correction for multiple testing was fairly uncommon. Therefore, the number of reported false-negative and false-positive results was likely to be high. Problems also arose from the use of heterogeneous phenotypes. An important intrinsic drawback of candidate gene association studies is the requirement of a hypothesis. Since knowledge about the pathophysiology of many diseases was incomplete, this type of study could at most identify a small amount of the relevant genetic variants. A final point of criticism concerns the fact that single genetic marker analysis is not a very genuine representation of how gene products act in vivo. The combination of several SNPs in a gene or even in a pathway and its on-aggregate testing for disease association might represent a more promising approach.

1.1.5 Identification of Genes: The Genome-Wide Approach

In the year 2005, GWAS appeared as a novel method to conduct genetic association analyses. In a GWAS, a set of hundreds of thousands of genetic polymorphisms representing the entire genome is put on a genotyping array. Originally, only common SNPs (those with a minor allele frequency, MAF > 5%) were considered, assuming that they would play the most important role in the underlying trait or disease architecture (common disease-common variant hypothesis). All the genotyped SNPs are then tested one after the other for disease association in a hypothesis-free manner. The GWAS approach makes use of the LD structure in the genome. If two SNPs are in LD, i.e. if they are inherited together more often than expected by chance, they may serve as proxies for each other. A set of less than one million SNPs could thus tag over 90% of the common genetic variation at the population level [14]. The international HapMap [15] and 1000 genomes [16] projects revealed the correlation structure by determining the haplotypes of all the sequenced individuals. The LD structure in the HapMap and 1000 genomes reference panels is also used to predict the genotypes of ungenotyped SNPs. Such an imputation procedure facilitates the comparison of data deriving from different genotyping platforms.

GWAS are typically carried out in a two-stage design to combine high power with reduced genotyping costs. A discovery sample is investigated for statistically significant associations, which are then tested in an often larger replication sample. SNPs are usually selected for replication by their P-value in the association (typically $P < 5*10^{-8}$, considering Bonferroni correction for one million independent tests) or by the control of the false discovery rate. Due to the winner's curse bias in the discovery sample, the expected effect in the replication set is smaller. Joint analyses of discovery and replication sample are also often reported in the literature since they are slightly superior in terms of statistical power [17]. The statistical power, i.e. the probability of detecting a truly associated variant, depends generally on the effect size and the frequency of this variant, the underlying disease model, the sample size and the SNP coverage of the array. An additive genetic model (in which risks increase additively per allele on the log-scale) is usually most plausible; and as it meanwhile turned out that most SNPs contributed individually only marginally to disease outcomes (typically conferring relative risks in the range of 1.1 to 1.3), optimal sample sizes reach several thousands.

GWAS have produced more robust results than candidate gene association studies for a number of reasons. Stringent correction for multiple testing is applied, which minimizes the risk for false positives. Confounding by ethnicity is normally taken into account by adjusting the regression models for principal components standing for differences in population substructure [18]. Quality control measures further reduce the chance of inaccurate SNP assessment. In the past years, GWAS have evolved to the most important tool to discover genetic variants influencing traits with complex aetiology. Over 1600 papers have been published reporting more than 2000 robust associations with more than 300 human traits and diseases [19]. The original concern that many top findings were outside of coding regions has been constantly weakened and replaced by the view that common SNPs lying in regulatory regions are more important with respect to disease associations than those lying in coding regions. Many novel mechanisms of disease aetiology have been reported, but limitations derive from the fact that only relatively frequent SNPs and those which tag the genome have been considered. In order to infer causality, more sophisticated GWAS and follow-up methods are indispensable (Table 1 and see 1.1.6, 1.1.7). This may also relieve the hitherto most strongly debated concern about the relevance of GWAS results, namely the fact that they only explain little of the estimated heritability of the respective traits or diseases (see 1.1.8, 1.1.9).

Table 1. Current methodological trends to complement and refine results from traditional genome-wide association analyses.

Progress within the GWAS approach	*Progress beyond the GWAS approach*
Large sample sizes and meta-analyses	Targeted fine-mapping and conditional analyses
Advanced genotyping arrays and reference panels	Transcriptomics and proteomics
Homogenous populations and trans-ethnical comparisons	Functional annotation (*in vivo*, *in vitro*, *in silico*)
Refined and intermediate phenotypes	Pathway association analyses
	Whole exome and whole genome sequencing

1.1.6   Identification of Genes: Where Does the Genome-Wide Approach Move to?

There are currently several observable trends to make the GWAS output more informative.

First, the number of included samples gets higher in order to gain statistical power. Since pooling of different data sets (joint analysis) is often difficult to achieve, large consortia emerged which combined GWAS results in meta-analyses by exchanging summary

statistics. This increases the power in a similar way as joint analyses would do, but without the need to share individual-level data. Moreover, study-specific sets of covariates are allowed, and heterogeneity between the studies (e.g. due to different genotyping arrays, imputation methods, ethnic groups or covariate assessment) can be assessed. GWAS meta-analyses became meanwhile the standard for investigating many traits and diseases. They were able to reduce the number of false-positive findings, detected robustly associated variants with steadily decreasing effect sizes and reduced some of the missing heritability [20]. Another advantage of such large-scale endeavours is the prevention of publication bias issues.

Second, the design of the genotyping arrays gets more advanced. Modern chips like the Illumina Omni5-Quad contain already five million SNPs and claim to capture variation in the full low-frequency range (MAF $1 - 5\%$) in European, African and Asian populations. Moreover, this chip covers more than 73% of all SNPs found in the sequenced individuals of the 1000 genomes project with high LD ($r^2 > 0.8$) [21]. By imputing the GWAS data with 1000 genomes reference panels, up to 30 million variants may get tested. Developments abandoning the hypothesis-free array design are currently also promoted. Chips are produced which collect SNPs lying in loci known to be important for certain traits (e.g. the MetaboChip [22]).

Third, comparisons of trans-ethnic GWAS results get more common. Under the assumption of same disease mechanisms among different ethnic groups, such comparisons help to strengthen the findings and could often refine the results due to the different underlying LD structure. High levels of diversity in the African genomes let research communities formerly focus on GWAS in Europeans and Asian populations. But newer reference panels take African populations more into consideration, allowing for better SNP coverage on the arrays and better imputation quality. Besides ethical reasons, the origin of human history, the higher prevalence of communicable diseases, but also the often larger differences in exposures and co-morbidities are good reasons not to let Africa aside in genetic epidemiology [23].

Forth, phenotypic heterogeneity is regarded as a problem in the comparability of GWAS on a particular outcome. There is a trend to use more refined phenotypes or disease subtypes in GWAS. The major drawback is the loss of statistical power, but there are methods with minimal power loss that divide the outcome into different subtypes allowing for different

genetic effects [24]. Another approach is the usage of intermediate phenotypes. This is promising since genetic variants do not directly act on higher-order diseases. Stronger results are therefore expected by testing the genomic variability against a proxy for the manifested disease either on the cell (e.g. a cellular transcript or protein pattern) or on the tissue level (e.g. the concentration of blood or urine metabolites [25]). A kind of a large-scale Mendelian randomisation approach recently showed that SNPs associated with metabolic traits could also be associated with medically relevant phenotypes, suggesting hence causal links between the metabolic traits and the phenotypes [26]. This could help to identify the crucial pathways in the aetiology of diseases. There is generally some evidence that a limited number of pathways, especially those for systemic inflammation, could be important for a large number of NCDs [27]. Some recent GWAS approaches did therefore not concentrate on one or a few separate metabolites, but created metabolic network outcomes by the correlation structure of many metabolites [28]. Such approaches may reveal promising loci playing a role in the aetiology of many NCDs.

1.1.7   Follow-up on GWAS: How to Detect the Causal Variants?

The most promising approach to refine a GWAS signal at a certain locus is targeted fine mapping, in which the respective locus is sequenced in a number of samples often followed by genotyping the detected variants in a larger sample. Conditional analyses in order to find independent effects within a locus can directly be applied if individual-level data are available. Even if only summary statistics for individual SNPs are available, conditional and joint effects can be estimated with the help of LD estimates, which can also be taken from public domain data [29]. Targeted fine mapping will most likely lose importance due to the meanwhile very high coverage of modern genotyping arrays in the common and low-frequency spectra. Targeted sequencing would however still reveal rare variants, which can be used on aggregate to assess excess in cases or extreme phenotypes.

Large-scale expression studies have recently become a popular follow-up method on GWAS. Expression quantitative trait loci (eQTL) are publicly available for a number of tissues, and look-ups in these data sets are often conducted to confer functionality to a SNP. An eQTL study is done by linking an expression trait with the most associated SNP. The SNP may lie within one megabase of the transcription start or stop of the corresponding gene (cis eQTL) or outside this range including on other chromosomes (trans eQTL) [30].

Non-association does not deduce that the SNP is irrelevant with respect to the gene product. Namely, the SNP could act on the protein structure without modifying the expression levels (e.g. the *SERPINA1* PiZ variant is not associated with the *SERPINA1* transcripts [28], but with the protein levels, see 1.4.2), or its impact on the expression could be restricted to a specific tissue cell type, time-window or to the presence of an environmental factor. Another way to assess a role for a SNP in influencing expression levels is by determining allele-specific expression. It is measured in individuals heterozygous for the candidate SNP by assessing via RNA-sequencing if allelic transcripts deviate from a one-to-one ratio [31]. A further step would be the association of SNPs with the protein products of a cell (protein-QTL). The accurate and reliable assessment of the whole proteome at a certain time point is still challenging though, but recent methodological progress towards this aim has been reported [32].

The actual strength of the GWAS design is also a weakness: prior knowledge is not taken into account. Bioinformatic tools are valuable instruments to estimate the likelihood of a GWAS signal to be causal. While prediction of the impact of exonic SNPs on protein structure is well-established, functional prediction of non-exonic SNPs is more challenging. However, recent rapid progress resulted in the fact that over 80% of the human genome is currently allocated to some biochemical function in at least one cell type according to ENCODE [33]. Predicted elements with functional relevance include promoters and enhancers (via DNA methylation or histone modification patterns), chromatin accessibility (via DNAse I footprints), copy number variations (CNVs), transcription factor binding sites (via chromatin-immunoprecipitation, ChIP) and microRNA binding sites. Synonymous variation in coding regions could also play a greater role in functionality than previously expected owing to translational efficiency changes [34]. The main challenge in predicting functionality comes from the tissue-, time- or environment-dependency of these elements.

Functional annotation by *in silico* methods cannot reliably predict *in vivo* relevance. Animal models therefore remain important, and so do *in vitro* cell cultures. Some of the challenges these experimental methods face are the low penetrance of the causal variants associated with complex diseases, differences in the genetic structure between animals and humans, the general lack of transferability between animals and humans and last but not least the lack of resources to carry out such studies for all GWAS signals.

Pathway association methods are another approach to include knowledge and functionally bridge genetic variation with a trait or disease. One could either compare pathways for enrichments of GWAS top signals or test if SNPs lying in the genes of a certain pathway are associated with the outcome. The burden of multiple testing can be substantially reduced by such methods, but several challenges apply. There are for instance various ways how to define a pathway (usually via publically available databases), how to allocate SNPs to a gene (e.g. by physical location or by eQTL data) and how to deal with different gene lengths or SNP densities in the association tests [35,36].

Finally, since sequencing costs rapidly decreased in recent years, whole exome sequencing (WES) and eventually whole genome sequencing (WGS) are expected to rapidly gain ground. In a first step, they may lead to better SNP arrays and reference panels for imputation, but eventually, they may generally replace genotyping procedures. However, compared to association tests with common or low-frequent variants, the methodology for association tests with rare variants (MAF < 1%) is far less standardized. Testing all rare variants separately would lead to a huge burden of multiple testing, and, unlike for common variants, relying on a much smaller set of proxies is not meaningful due to the much lower level of correlations. Since rare variants are often population-specific, accounting for ancestry will be most crucial to draw valid conclusions from association results. In addition, association studies with rare variants have low power even when the sample is large. The most common strategy applied is the collapsing of rare variants in the genetic region of interest. Such methods have been published, and their performance in simulated data sets depends strongly on the number of assumed non-causal variants as well as on the presence of effects of different magnitudes and directions in the region of interest. They differ also by the possibility of accommodating covariate information or including interactions [37]. The allele-specific allocation of rare variants in a certain region of interest would also be a desirable property before combining them, but it is not straightforward to derive haplotype information from sequence or genotype data. It is important to note that, independent of the unsolved methodological challenges, aggregate rare variant analyses always lead to an underestimation of the contribution of rare variants to complex traits [38].

1.1.8 Where is the Missing Heritability? Non-Additive Components and Gene-Environment Interactions

There is considerable debate why GWAS results for most traits and diseases only explain a small part of the estimated heritability. One hypothesis propagates the erroneous estimation of heritabilities by pedigree-based methods owing to an underestimation of shared environment effects [39]. Current narrow-sense heritability estimates may further be inflated by the fact that they are based on the assumption that non-additive components (see 1.1.3) represent a negligible fraction of the broad-sense heritability. While dominance effects are indeed not believed to be of high importance to the missing heritability problem, it was recently reported that epistasis could explain a substantial part of current heritability estimates [40].

The presence of a gene-gene interaction is usually assessed by evaluating if the combined risk of the two genetic risk factors departs from multiplying the relative risks of the separate risk factors. Gene-gene interactions seem biologically highly plausible, but established examples in humans have not been reported. Relevant contributions of epistasis to many outcomes were observed in yeast [41]. Since yeast is a haploid organism which can be held under uniform environment, epistasis as the only non-additive component contributing to heritability could be easily estimated. In humans, the computational burden of testing epistasis on a genome-wide scale would be enormous due to the sheer number of possible interactions. A reduction in the number of tests could be achieved by only focusing on the SNPs with the strongest main effects or on those potentially related to known protein-protein interactions. Two-stage design methods with reduced computational burden and only minimal power loss have also been proposed [42].

Gene-environment interactions are assessed in an analogous way. They are often not considered in heritability studies since they are difficult to estimate. Moreover, erroneously ignoring them would rather inflate the environmental component than the heritability estimates [5]. There is however some disagreement in the literature if their presence could actually reduce the missing heritability problem. In any case, the assessment of such interactions is crucial in order to identify those individuals who are most susceptible to certain environmental influences. Support for the existence of gene-environment interactions comes from animal studies, in which the environment can be modelled and such interactions therefore more easily detected [43].

In humans, the presence of such interactions also seems likely. Environmental factors may only exert their effects in the presence of certain enzyme variants (e.g. the effects of certain toxins may depend on the genetic variation in the detoxifying enzymes). Analogous to the history of genetic association studies, the primary approach to assess gene-environment interactions were candidate gene-environment interaction studies. Although this term describes different concepts, there is always an underlying biological plausible hypothesis [44]. The investigation of a genetic variant modifying a known environmental cause of a disease represents a first concept. The candidate gene would be selected by lying in the pathway in which the environmental effect is believed to work. A reverse perspective is taken when we test the role of an environmental factor which potentially acts on the known genetic effect causing the outcome. Even if the genetic effect was only assumed to be causal (e.g. if the genetic variant was associated with the outcome in a GWAS in the absence of any functional knowledge), testing for such an interaction would be a valid candidate gene-environment interaction approach. A positive result would in that case confer function to the functionally unknown genetic variant. A much younger approach based on the GWAS concept are genome-wide interaction studies (GWIS), which do not imply a hypothesis [45]. Studies with genome-wide as well as environmental data can test each single SNP for interaction with an environmental factor in the same way as performed in a GWAS.

In spite of considerable efforts, very few successfully replicated examples of gene-environment interactions have been reported in humans [46]. This is mostly attributed to methodological challenges. They are present in the exposure assessment (e.g. methods, exposure distributions and measurement errors may vary between different studies), in the candidate design (e.g. publication bias) and especially in the lack of statistical power. It is estimated that a four times larger sample is needed to detect an interaction compared to a main effect of similar magnitude [47]. Like in any genome-wide approach, the high burden of multiple testing is also a major issue in GWIS. Several multi-step designs for GWIS have been proposed to combine cost efficiency and statistical power [45].

There is a last aspect on a more philosophical side why it is so difficult to observe robust gene-environment (or gene-gene) interactions. It is argued that statistical interaction as described above is not very adequate to describe biological interaction [48]. Namely, the same statistical model is applied to evaluate whether effects of one factor over different levels of the other factor go in opposite directions or in the same direction, but at different

strength. The model is furthermore supposed to correctly evaluate if a factor depends on the presence or absence of another factor like assumed in cellular pathways. It is therefore questionable if simply more accurate data and larger sample sizes would lead to compelling evidence of gene-environment interactions. More sophisticated modelling techniques for interactions could prove useful, but epidemiologic observations might be inherently limited to infer mechanisms of biologic interactions [49]. Some authors see non-parametric methods like data mining and machine learning, which include prior knowledge, as much more promising to detect biological interactions [50].

1.1.9 Where is the Missing Heritability? Additive Components

Most researchers still see the bulk of the missing heritability in additive components. More heritability might be hidden in common variants because GWAS are still underpowered to detect such variants if they have very small effect sizes. Hence, many more common variants significantly associated with the respective outcome would be expected by simply enlarging the sample size [51]. Others see rarer variants not covered by current chip designs as the culprit for the missing heritability problem. Larger application of WGS in association analyses will soon provide answers on that issue. CNVs are also neglected on genotyping arrays, and there is some evidence that they not only play an important role in causing rare Mendelian diseases, but also impact common complex diseases [52].

More challenging is the estimation of the impact of epigenetics. Although epigenetic modifications like histone alterations and DNA methylations generally question the concept that genetic variation is stable over time and in each cell from the start of conception, they may be of minor importance to heritability estimates since they are mostly reset during gametogenesis and evolve newly during development. However, some modifications are most likely transmissible from one generation to the next. Inheritance models of the environment-sensitive epigenetic variation showed that a relevant contribution to the missing heritability issue was possible [53]. Several rare genetic disorders are in fact influenced by genomic imprinting [54], an epigenetic mechanism leading to the expression of only one allele.

1.2   Asthma and Chronic Obstructive Pulmonary Disease

Asthma is a very heterogeneous disease, typically defined by reversible airflow obstruction in combination with wheeze, cough and shortness of breath. The inflammation of the airway wall is mainly driven by different subsets of T helper cells and by eosinophils. The associated airway remodelling process results in hypertrophy and -plasie of bronchial smooth muscle tissue, thickening of the airway mucosa as well as stronger mucus production and blood supply in the airways [55]. Asthma is often, but not always, accompanied by a general tendency to be allergic and can therefore be divided in atopic (measured by a positive skin prick test or generally high immunoglobulin E levels) and non-atopic. Another distinction is early-onset (before puberty) vs. late-onset asthma. Obesity-related asthma has recently also been described as a specific phenotype [56].

In chronic obstructive pulmonary disease (COPD), a term which combines chronic bronchitis and emphysema, the airflow obstruction is not fully reversible. The underlying inflammatory process is predominantly based on neutrophils and cytotoxic T cells and results, together with oxidative stress and elevated proteolytic digestion in the lung parenchyma, in fibrotic changes and narrowing of the small airways as well as in loss of tissue walls between the alveoli [57]. Symptoms include chronic cough, wheezing, phlegm and shortness of breath. There is obviously a high overlap between the two diseases, but as opposed to asthma, the inflammation in COPD includes the lung parenchyma, is less prominent with eosinophils, is more resistant to corticosteroid treatment and usually gets worse over time [58]. A clear distinction between the two diseases is not always possible though, in particular since the asthma-COPD overlap syndrome, i.e. the coexistence of both diseases, is phenotypically heterogeneous and relatively common, especially in the elderly [59,60].

Both asthma and COPD are typical NCDs with chronic characteristics and complex disease aetiologies. Environmental exposures are strong determinants for disease manifestation, but susceptibility is assumed to depend on many genes. The genetic contribution is believed to be high due to the fact that equal exposures lead to high individual differences in terms of disease susceptibility [61]. In fact, heritability estimates for both asthma and COPD lie at around 60% [62,63].

1.2.1 Public Health Burden and Diagnosis

It is estimated that up to 300 million people worldwide suffer from asthma and the annual death toll is 0.25 million [64]. Corresponding World Health Organisation estimates for COPD are 64 and 3 million. While COPD is ranked in the top ten in terms of DALYs over all age-groups, asthma is very prominently ranked during childhood [1]. The relevance of both disorders is expected to rise with Western lifestyle and globally still increasing smoking rates [65].

Asthma is usually diagnosed by the enquiry of symptoms and by the assessment of variable airway obstruction. This is done by comparisons of pre- and post-bronchodilation spirometry and by assessment of responses to bronchial provocation (i.e. by testing airway hyperresponsiveness). As this is a very laborious and stressful procedure, common practice in epidemiological studies is to use questions about an asthma diagnosis by a physician or about prescribed asthma medication. There is evidence that such information represents a valid proxy for real asthma [66].

COPD is a very heterogeneous syndrome and clinical presentation, physiology, imaging by computed tomography scanning, decline in lung function or response to therapy differ substantially. Although this insufficiently captures the heterogeneity of the disease [67], COPD is mostly assessed by measuring the degree of airflow obstruction by spirometry after bronchodilation. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) recommends as diagnostic criteria, in addition to measuring the severity of airflow obstruction, the assessment of symptoms, the history of exacerbations and the presence of comorbidities [68]. The spirometric definition of COPD is $FEV_1/FVC < 0.7$ (FEV_1 stands for the air volume exhaled during the first second of forced expiration and FVC for the maximally exhaled volume, see 1.3.1). Four grades of severity (from mild to very severe) are defined according to the correspondent FEV_1 value, which is set in relation to a predicted value (%-predicted) depending on the individual's sex, age, height and a regional context. It is widely acknowledged that the mentioned cut-off value for the FEV_1/FVC ratio is a suboptimal solution. Since this ratio naturally declines with age, elderly people are overrepresented in COPD cases. This misclassification issue has brought up the proposal of using the lower limit of normal of the FEV_1/FVC ratio to define the threshold of obstruction. It is calculated as the fifth percentile of the normal distribution in healthy never-smokers for a given population according to sex, age and height.

1.2.2 Risk Factors for Asthma

Asthma prevalence shows dependency on sex with lower rates in girls than boys, but consistently higher rates in adult women than men [69]. Sex hormones seem to play a role in this sexual dimorphism as postmenopausal women taking hormone replacement therapy show higher rates than those not taking them [70]. Black Americans and Hispanics seem to have higher rates than White Americans [71]. Western lifestyle in general seems a key driver for the observed rise of asthma [72], and obesity is associated with asthma in numerous studies [73]. Further risk factors associated with asthma include maternal smoking during pregnancy [74], early exposure to tobacco smoke, infections in early life, exposure to house-dust mites or pets as well as indoor and outdoor air pollution [72].

Before the first efforts to carry out GWAS started, some 120 candidate genes had been described in the literature to be associated with asthma or related phenotypes, including a handful genes determined by genetic linkage studies and subsequent positional cloning [75]. The genes represented categories like innate immunity, T helper cell differentiation, lung function or airway remodelling. Due to the typical limitations of candidate gene association studies (see 1.1.4), reviews were only conducted in a non-systematic style, and firm conclusions could not be drawn from these results.

The first GWAS on (childhood-onset) asthma pointed to *ORMDL3* [76], a signal later refined to *GSDMB*, for which interaction with environmental tobacco smoke (ETS) was also observed [77]. Candidate genes previously associated with asthma were looked up in the results of a GWAS on early-onset asthma, and only four genes (*TGFB1*, *IL1RL1*, *IL18R1* and *DPP10*) remained plausible candidates [78]. Such investigations further discredited the design of candidate gene association studies, but did not directly argue against their validity, as GWAS ignored environmental factors. Larger GWAS as well as GWAS meta-analyses confirmed the *GSDMB* locus and extended the number of statistically significant findings especially by different interleukin receptor-related polymorphisms [79-81]. This may highlight the importance of immunoregulation in the pathogenesis of asthma. While the first GWAS of asthma in Africans failed to confirm any hitherto known loci [82], results in ethnic strata in later GWAS pointed to considerable agreement in genetic effects across the ethnicities [81,83]. GWAS on discriminable asthma phenotypes is likely to be a promising strategy to refine current signals [84]. Currently, more than 30 GWAS on asthma are listed and more than 50 SNPs have been linked to asthma [19].

Further efforts to find more loci or to refine the detected signals have been carried out. Fine mapping procedures in asthma candidate genes led to the discovery of asthma-associated rare variants [85]. Expression analyses of GWAS top hits were sometimes conducted [76], and studies prioritizing GWAS findings with eQTL results revealed a few further signals and important networks in asthma pathogenesis [86,87]. Evidence for some of the interleukin-related genes came also from functional studies [88,89]. Asthma-associated non-coding SNPs have recently been shown to be enriched in genomic regions, acting as promoters or enhancers in immune cells contributing to asthma [90]. Pathway analyses resulted in the identification of 21 potentially causal pathways [91]. In terms of interactions, mostly genes involved in immune pathways have been tested in combination with smoking, air pollution, microbial or workplace exposures [92,93]. Owing to the same reasons as in candidate gene association studies, and additionally due to the often very crude exposure assessment, it has been difficult to draw firm conclusions from these reviews. A first GWIS on asthma and atopy found no statistically significant interaction between a genetic polymorphism and farming environment [94]. There was also no strong evidence for genetic interactions with smoking or hay fever status in a genome-wide scan [95].

Taking together, several environmental and genetic factors have been observed to be robustly associated with asthma. However, the major part of the heritability remains unexplained by current knowledge, and very little is known how the environment and the genetics work together.

1.2.3 Risk Factors for COPD

COPD occurs much more in men than in women, but this is most likely due to the different smoking habits. Although COPD can generally be described as a smoking disease [96], up to 45% of concerned patients have never smoked [97]. Underestimated for a long time, smoke from biomass fuel for heating and cooking (indoor air pollution) is possibly at least as important since the number of exposed people is believed to be higher than for smoking. Other inhalation exposures such as outdoor air pollution from industry and traffic, occupational hazards as well as respiratory infections including tuberculosis are also of relevance [97-99]. Systemic inflammation is associated with COPD outcomes, but a large Mendelian randomisation study pointed against causality [100].

On the molecular level, the state of two equilibrium conditions seems crucial for the COPD pathogenesis. The protease-antiprotease hypothesis is mainly driven by the early observation that genetic polymorphisms leading to a severe lack of alpha1-antitrypsin (AAT) were associated with early-onset emphysema (see 1.4.3). The oxidant-antioxidant hypothesis is based on the observation that cigarette smoke contains high concentrations of oxidants. The two hypotheses are physiologically linked via inflammatory pathways. Not surprisingly, predominantly variants in genes allocated to these pathways have so far been selected as candidate genes in COPD association studies. Examples include the genes for matrix metalloproteases, heme-oxygenases, glutathion S-transferases, transforming growth factor beta and tumour necrosis factor alpha. Replication studies produced often inconsistent results, and meta-analyses pointed to very few robust associations [101,102]. Genome-wide linkage analysis of lung function measures and airway obstruction in severe COPD was a more unbiased way to suggest candidate genes and found evidence for regions on chromosomes 2q and 12p [103,104]. These regions could later be narrowed down to the genes *SERPINE2* and *SOX5* in independent candidate gene association studies on COPD [105,106].

A first GWAS on COPD detected *CHRNA 3/5*, a locus previously associated with smoking behaviour [107]. Since this locus was not associated with the amount of smoking in the GWAS, its effect might at least be partially independent of smoking behaviour. *HHIP*, a locus simultaneously described by its association with lung function [108], was also associated with COPD. A subsequent GWAS confirmed these results and found *FAM13A* (also associated with lung function [109]) as a new signal [110]. It was later shown that these three loci affect different subphenotypes of COPD [111]. The importance of a detailed phenotypic characterization in COPD studies was further taken into account when a GWAS on computed tomography-assessed emphysema was carried out and identified *BICD1*, a gene associated with telomere length (cell aging), as the only genome-wide associated locus [112]. More recently, larger GWAS on COPD as well as on lower limit of normal-based airflow obstruction found evidence for a locus on chromosome 19q13, previously associated with smoking behaviour, and for *HTR4*, a gene previously related to FEV_1/FVC [113,114]. In general, many regions associated with lung function, typically identified in much larger GWAS (see 1.3.2), show nominally significant association with COPD [114,115].

Further steps to determine the causal variants have been reported. Expression of the *CHRNA 3/5* locus in lung tissue [114] and a regulating mechanism for *HHIP* expression were published [116]. *SOX5* was shown to be necessary for normal lung development in mice [106]. Epigenetic patterns in numerous genes which lie in biologically plausible pathways were found associated with COPD, making this a promising biomarker [117]. However, reverse causality could also explain these associations, and some may question in general the usefulness of the methylome in peripheral blood cells with respect to lung diseases. Although COPD represents the classic paradigm of gene-environment interactions (see 1.4.4), very few interaction studies have been conducted. In one of the largest candidate gene association studies, the *MMP12* association with COPD was restricted to smokers [118]. Some evidence for the presence of interactions of ambient air pollution with genes in the oxidative stress system was reported [119]. No study on COPD is currently available in the literature that assessed gene-environment interactions on a genome-wide scale.

Taking together, knowledge on environmental risk factors is better for COPD than for asthma, but the genetics are even less understood. There are fewer and smaller GWAS on COPD or airway obstruction available, probably owing to the fact that asthma-diagnosis often relies on simple self-reports. Nevertheless, the high overlap with the genetics of general lung function is a promising feature for a better understanding of the COPD genetics in the future.

1.3 Lung Function

Lung function is the most widely used marker to assess airflow obstruction (see 1.2.1). It is furthermore inversely associated with higher cardiovascular risk and general mortality [120,121]. Deficits can arise from diminished growth during childhood and adolescence or from accelerated decline during adulthood.

1.3.1 Measures

The spirometric testing procedures must fulfil quality standards [122]. These include for instance guidelines for device maintenance and patient instructions. The motivation of patients to invest the largest possible effort in the manoeuvres is thereby crucial. Most frequently assessed measures are FEV_1 and FVC. Its ratio is a measure of airway

obstruction as narrowing of the larger bronchi primarily affects FEV_1 and to a lesser extent FVC. The forced expiratory flow during the middle half of FVC ($FEF_{25-75\%}$) is an alternative measure, which correlates generally well with FEV_1, but which shows a higher variability. The middle half of the FVC volume is thereby divided by the time of duration for its expiration. Changes in $FEF_{25-75\%}$ in airflow-obstructed patients are often more pronounced than those in FEV_1, and it was suggested that $FEF_{25-75\%}$ can provide additional evidence to FEV_1/FVC for small airway disease [123].

1.3.2 Risk Factors

Environmental risk factors for lung function are generally those already listed for COPD. Heritability estimates for lung function measures were shown to be somewhat lower than for the disease outcomes and varied between 0.25 and 0.45 for the general population as well as for individuals with impaired lung function [124,125]. The heritability for longitudinal lung function is probably lower [126].

Candidate gene association studies looking at cross-sectional or longitudinal lung function in the general population have not been as frequently carried out as those in patient cohorts or as those using a disease-related outcome definition. A genetic linkage study identified a region on chromosome 6q27 as a candidate locus [127], a signal later assigned to *SMOC2* [128]. The first large GWAS on lung function used %-predicted FEV_1/FVC as outcome and identified *HHIP*, a locus contributing to lung morphogenesis, as an associated locus [108]. Large GWAS meta-analyses by the SpiroMeta [129] and Charge [109] consortia confirmed that locus and found several others associated with FEV_1 and FEV_1/FVC, many of them biologically plausible candidate genes in lung remodelling or morphogenesis processes. Interestingly, a thorough analysis focusing on previously defined candidate genes for COPD and lung function could not identify any region associated with lung function in the results of the GWAS meta-analysis by SpiroMeta [130]. By further increasing the sample size to almost 50 000 individuals in the discovery set and a similar number in the replication set, 16 further regions were observed as risk loci, and expression in lung tissue for most of the transcripts was reported [131]. Those top-ranking SNPs did not interact with smoking. A recent meta-analysis on lung function in a similar study sample jointly considered gene main effects and gene-smoking interactions on a genome-wide scale [132]. This resulted in

three novel loci of potential importance for lung function, although interaction analyses per se did not yield any significant results on the genome-wide level.

Studies on longitudinal lung function have so far been carried out in much smaller samples and no significant results on the genome-wide level have been reported. In one GWAS meta-analysis, suggestive loci showed a strong heterogeneity according to asthma status [133]. In another study, associations of *TMEM26* and *FOXA1* with lung function decline were identified in individuals with mild COPD, but these findings were not successfully replicated in general population cohorts [134].

We summarize that, despite the higher sample sizes in the GWAS and a much more homogeneous phenotype as opposed to asthma or COPD, the heritability for cross-sectional lung function explained by the associated SNPs is still very low and has so far not exceeded 10% [131].

1.4 Alpha1-Antitrypsin

AAT is a 52-kDa glycoprotein and the most prevalent protease inhibitor in the human blood with serum concentrations of about 120 mg/dl. It is mainly synthesized in the liver, and its main function is the protection of lung parenchyma from digestion by neutrophil elastase (Figure 1). Deficient blood levels were first detected by unusual serum protein electrophoresis patterns some 50 years ago [135].
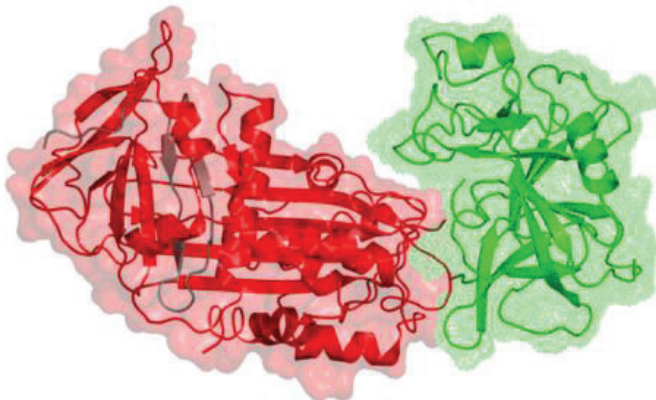

Figure 1. Structure of a complex between AAT (red) and elastase-1 (green). *Source: Protein Data Bank in Europe (PDBe).*

1.4.1 Environmental Determinants

Being an acute-phase protein with anti-inflammatory properties, the concentration of AAT can rise up to four-fold of normal and stays elevated for a week in case of an acute inflammation. Sex, age, blood pressure, smoking and body mass index (BMI) were all associated with AAT, independent of the presence of an inflammatory condition [136]. Elevated AAT levels were also measured in individuals with lung function impairment, but adjustment for the inflammatory marker C-reactive protein (CRP) cancelled out this association [136].

1.4.2 Genetic Determinants

The heritability of AAT has to our knowledge never been estimated, but may lie in the range of other serum proteins (e.g. 0.3 – 0.4 for CRP and albumin in middle-aged White Americans [137] or 0.5 – 0.6 for a wide range of serum lipid proteins in a population of young Finnish adults [25]). The protease inhibitor S and Z variants (PiS and PiZ) are the most common AAT deficiency alleles. The alphabetic designation is based on the mobility of the protein form in an electric field as used in isoelectric focussing (phenotyping). The M variant (PiM) is the variant with normal mobility properties. Phenotyping has long been the standard to determine deficiency alleles, and only recently, genotyping by real-time fluorescent polymerase chain reaction (PCR) has started to gain ground. The PiS and PiZ variants are non-synonymous SNPs in *SERPINA1*, the gene encoding for AAT. They do not influence AAT expression levels, but lead to a reduced secretion of plasma protein in the liver due to intracellular erroneous processes at the protein level. While the PiS form results in an increased degradation [138], the misfolded PiZ form accumulates and polymerizes in the hepatocytes [139], leading to chronic apoptosis and a higher risk for neonatal hepatitis [140] and cirrhosis or liver cancer later in life [141]. While individuals homozygous for PiS still produce approximately 70% of normal AAT levels, PiZZ carriers only exhibit about 15% of normal serum levels [142]. Highest prevalence occur on the Iberian Peninsula for PiS (up to one homozygote in 500 persons) and in Scandinavia for PiZ (up to one homozygote in 1500 persons) [143]. Dozens of rarer deficiency alleles have been described in case reports, and a few null alleles (PiQ0), in whom no secretable protein is built, are also known [144].

1.4.3   AAT Deficiency, COPD and Lung Function

PiZZ carriers are at higher risk for COPD, in particular shown for early-onset panacinar emphysema. This has led to the formulation of the protease-antiprotease imbalance hypothesis in the causation of COPD, meaning that reduced levels of AAT cannot sufficiently inhibit neutrophil elastase activity in the lung tissue. Support for this hypothesis comes from individuals with PiQ0 alleles, who also show elevated risks for COPD. The protective threshold, below which the association with COPD has been well-established, corresponds to about 40% of normal AAT concentrations (11 µM or 50 mg/dl). Such a strongly reduced AAT level, also known as severe AAT deficiency, is rare (mainly caused by the PiZZ genotype) and accounts for about 1 – 2% of all COPD cases. However, studies investigating the impact of PiZZ are hampered by ascertainment bias because they usually include disproportionally many people with lung disease (and their families). In fact, non-smoking PiZZ carriers identified by screening do likely not have a much increased mortality compared with the general population [145]. This lack of pathological symptoms is one of the main reasons why PiZZ is a strongly underrecognised genetic disorder. Determinants of airflow obstruction or elevated lung function decline in PiZZ individuals are, in addition to cigarette smoking, which is the by far most crucial factor, the patient's sex, age and asthma status as well as occupational exposure [146,147] and air pollution [148,149]. The heritability of lung function in persons with severe AAT deficiency is estimated at about 15% [150].

The PiS allele is only believed to be a risk factor for COPD in combination with the PiZ allele [151]. Individuals with such a compound heterozygosity (PiSZ) exhibit AAT serum levels around or slightly above the protective threshold. The impact of one PiZ allele (PiMZ), a condition which preserves about 70% of the normal serum level and is often referred to as intermediate deficiency, is controversially discussed. Case-control, cross-sectional and family-based studies slightly favour an enhanced risk for airflow obstruction [152,153]. The question if PiMZ individuals have a higher age-related lung function decline has also been addressed. While the answer may be more on the negative side for the general population [154,155], studies showed that in rapid decliners and incident COPD cases, the PiMZ genotype is disproportionally frequent, suggesting other predisposing factors [156-158]. Some of the rare deficiency alleles have also been associated with a higher COPD risk [144], but they do not play an important role at the population level.

Weekly intravenous doses of purified human AAT is the only remedy for individuals with severe AAT deficiency to hold levels constantly above the protective threshold. Its prescription is controversial though due to the non-proven efficacy and the high costs of treatment (well above US$ 10 000 per year). While it is widely used in PiZZ carriers in the USA, many European countries do not even offer this therapy. Meta-analysed results from available studies tend to give arguments in favour of the augmentation therapy [159], but the two proper RCTs were not supportive of a beneficial effect [160,161]. RCTs are, however, a suboptimal study design for a very slowly progressing disease like emphysema. There are a few ideas why substitution therapy may not be as efficacious as expected [144]. First, although the administered dose usually results in blood concentrations above the protective threshold, acute or chronic inflammatory conditions may require further up-regulated circulating levels. Second, AAT is also produced in alveolar macrophages as well as in bronchial and alveolar epithelial cells. Locally produced PiZZ polymers have been found at these sites [162], and their chemotactic, pro-inflammatory properties may maintain the enhanced level of tissue destruction [163]. It is therefore hypothesized that individuals with severe deficiency variants that do not polymerise would benefit most, whereas carriers of deficiency variants that build polymers (carriers of PiZZ and a few very rare genotypes) may not only have the enhanced risk of liver disease, but may also benefit less from augmentation therapy in terms of their pulmonary health.

1.4.4   A Textbook Example for Gene-Environment Interaction?

PiZZ carriers who smoke develop earlier and more severe COPD than non-smoking subjects with PiZZ genotype [164], and their FEV_1 decline is higher [165]. The interaction between deficiency variants in the *SERPINA1* gene and smoking has been considered the classic example of gene-environment interaction for several decades. The use of the term gene-environment interaction in that case was based on the idea that genes and environment work together to cause a certain disease. Gene-environment interaction in a statistical sense is a different concept (see 1.1.8) and was reported for the PiZ deficiency allele and smoking only once (with respect to cross-sectional FEV_1) [166]. Support for such a statistical interaction comes especially from the biochemical side. While the increased oxidative stress level in the lung constitutes the main effect of cigarette smoke, it may also directly affect the protease-antiprotease disequilibrium. Cigarette smoke oxidises and thereby inactivates

functional AAT [167]. Moreover, oxidation has also been related to AAT polymerization in mice [168], which may attract in return more neutrophils and proteolytic enzymes. Figure 2 summarizes the current understanding of how PiZZ genotype and cigarette smoke jointly exert their adverse impact on pulmonary health.
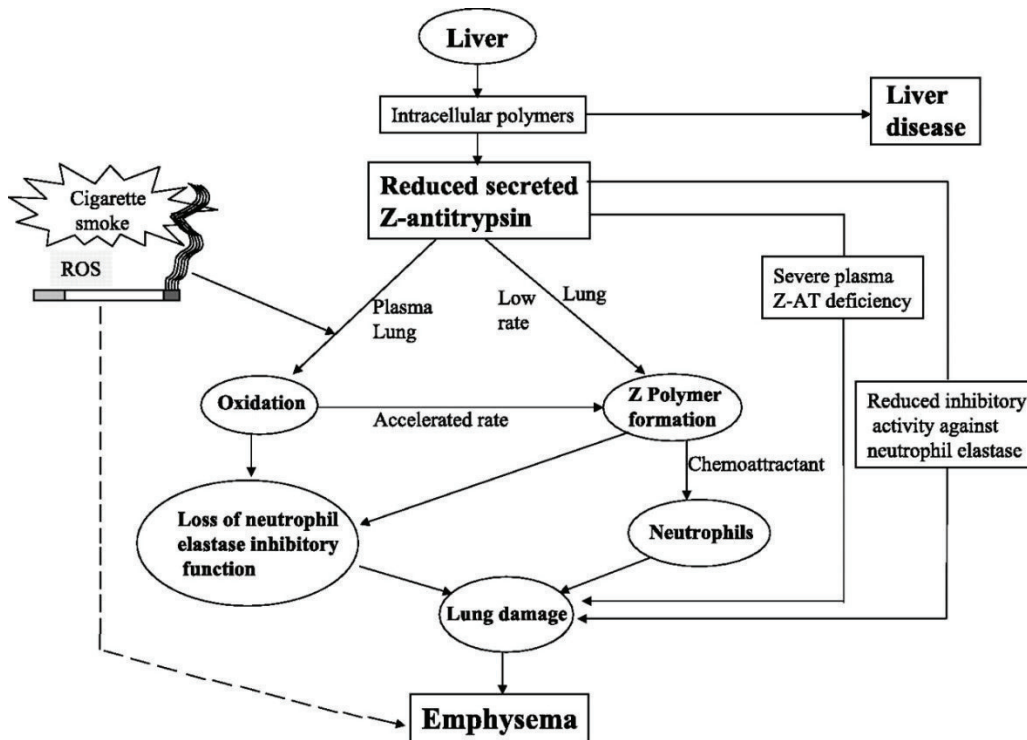

Figure 2. The role of the PiZ-form of AAT in combination with cigarette smoke in the development of emphysema. ROS, reactive oxygen species. *Source: Alam S, et al. (2011)*, [168].

In summary, there is severe lack of functional AAT in the lung because of a reduced secretion from the liver, a reduced inhibitory activity of the (non-polymerized) PiZ form and the local building of inactive PiZ polymers. This condition is exacerbated by the smoking-induced oxidation and additional polymerization of the remaining functional AAT, which deteriorates the influx of neutrophils, resulting in inflammation, proteolysis and finally tissue destruction.

Some studies looked at risk modification of AAT deficiency by occupational and environmental inhalants other than smoking. Unfortunately, a formal test for interaction has not been assessed in any study with adults, mostly due to the lack of investigating a control group [169].

2   Aims

2.1   Genetic Modifiers of the Obesity-Asthma Association

Obesity is a well-established risk factor for asthma in children and adults (see 1.2.2), but the underlying aetiology remains obscure. Obesity is associated with low-grade systemic inflammation, which may provide a link to the airway remodelling processes (including the altered survival of different airway cell types) observed in asthmatics. We investigated therefore if genetic polymorphisms believed to be important for cell division modified the obesity-asthma association. This is an example of a candidate gene-environment interaction study with the hypothesis that a known environmental cause is modified by a genetic variant lying in a pathway in which the environmental effect is believed to work (see 1.1.8). The selected SNPs had previously been associated with increased risk for several types of cancer [170,171] and had been suggested to modify the effect of air pollution change on longitudinal lung function [172].

Results are presented in chapter 4.1.

2.2   *SERPINA1* Deficiency Alleles and Reference Values in the General Population

Severely reduced AAT serum levels are responsible for an estimated 1 – 2% of COPD cases (see 1.4.3). Moreover, some *SERPINA1* genotypes leading to a severe deficiency are also associated with liver disease. Intermediate AAT deficiency (mainly caused by PiMZ genotype) is a risk factor in subgroups, but since this genotype is much more frequent than PiZZ, it may account for as many COPD cases as severe AAT deficiency [158]. The identification of AAT deficiencies is therefore crucial, and a first step towards this aim is the determination of AAT serum concentrations in dedicated laboratories. This diagnostic procedure is recommended for all patients with COPD, unexplained liver disease, asthma that is incompletely reversible after treatment with bronchodilators as well as for siblings of severely AAT deficient individuals [173]. It is thereby important to specify appropriate reference values for deficiency genotypes and to determine cut-offs below which further testing (in the form of phenotyping, genotyping or sequencing) is meaningful from an economical and public health perspective. Previously determined reference values were

assessed in patient study samples and hampered by either the application of outdates methods (e.g. radial immunodiffusion) or rarely used protein purification procedures and concentration units [142]. We aimed at establishing a strong phenotype-genotype correlation between AAT blood levels and PiM, PiS and PiZ alleles that also included the consideration of inflammatory conditions in a large population-based sample.

Results are presented in chapter 5.1.

2.3   Genetic Determinants of AAT Serum Level

The strong influence of PiS and PiZ alleles on AAT serum level dominates our knowledge about genetic variants influencing AAT (see 1.4.2). While several other rare deficiency or even null-variants in the *SERPINA1* gene are also present in the literature [144], common variants modulating AAT expression or protein secretion have not been reported so far, neither within the *SERPINA1* gene nor outside. We therefore performed a GWAS on AAT serum level in a population-based sample. Depending on the number and location of the top-ranking SNPs and their magnitude of association with AAT, replication in independent study samples, targeted fine-mapping or functional characterisation was conducted. AAT-associated SNPs were furthermore tested for association with lung function.

Results are presented in chapter 5.2.

2.4   *SERPINA1* PiMZ Genotype and Elevated Lung Function Decline: Which are the Predisposing Factors?

Heterozygous carriers of the PiZ allele are disproportionally frequent in COPD cohorts or among individuals with rapid lung function decline (see 1.4.3). Furthermore, within subgroups with clinically established COPD diagnosis, persons with PiMZ genotypes showed lower lung function than those with PiMM genotypes [174]. That points to a subset of PiMZ carriers which is at elevated risk for losing lung function or getting COPD when other predisposing factors are present. Another interesting observation suggesting further predisposing factors is the worse lung function in PiMZ parents of PiZZ individuals with COPD as opposed to the respective parents of PiZZ carriers without COPD, a difference which could not be explained by the amount of smoking in the past [175].

2.4.1 Inflammatory Triggers as Predisposing Factors

PiMZ individuals were reported to have worse pulmonary function when smoking than persons without the PiZ allele, whereas no difference was found in non-smokers [176]. Moreover, cigarette smoke is believed to act on AAT directly and therefore possibly interact with reduced AAT levels (see 1.4.4). Hence, we investigated if the lung function decline effect of one PiZ allele is dependent on the exposure to cigarette smoke or to other pro-inflammatory triggers like obesity. Incidence rates of respiratory health symptoms were also considered.

Results are presented in chapter 5.3.

2.4.2 Air Pollution and Occupational Exposure as Predisposing Factors

A few underpowered studies suspected interaction between occupational exposure or air pollution and AAT deficiency on longitudinal pulmonary health [169,177,178]. We aimed at elucidating more rigorously the role of the PiMZ genotype in the associations of lung function decline with occupational exposure to vapours, gas, dusts and fumes and with the change in ambient air pollution, respectively. This is of high public health interest because of the reduced capacity of individuals to avoid these types of exposure. We pursued in this project a reverse perspective of a candidate gene-environment interaction than in project described under 2.4.1 (see 1.1.8). While a possible environmental modification of the unevenly distributed genetic effect of the PiMZ genotype on longitudinal pulmonary health was examined in the previous case, we tested here a possible role of the PiMZ genotype on established associations between environmental factors and lung function decline. This latter approach is more useful with respect to targeting modifiable factors for prevention.

Results are presented in chapter 5.4.

2.5 Further Candidate SNPs Influencing the Beneficial Effects of Air Pollution Decline

Ambient air pollution is thought to be a risk factor for COPD (see 1.2.3). In particular, improved air quality with respect to particulate matter < 10 μm (PM_{10}) was shown to result in attenuated lung function decline [179]. Profit from air pollution decline in the general

population is likely dependent on genetic factors. The candidate genes tested for interaction with air pollution on pulmonary health have so far been selected by biological knowledge, i.e. by lying in promising pathways [119,172]. In this study, we selected SNPs robustly associated with lung function, but for which the determining mechanisms had not been elucidated [131] (see 1.3.2). By testing if these SNPs modify the association between improved air quality and reduced loss in lung function, we not only aimed to gain knowledge about genetic factors determining the differing profits from air pollution decline in the general population, but also to get insight into the functionally unexplained associations of these variants with lung function.

Results are presented in chapter 6.1.

3   Methods

All the different projects were carried out using the Swiss Study on Air Pollution and Lung Disease in Adults (SAPALDIA). A very brief overview about the study design and the data is given here; more detailed and project-specific methods can be found within the papers (chapters 4-6).

3.1   Description of SAPALDIA

3.1.1   Study Design

SAPALDIA is a cohort study which was initiated in 1991 in order to study the health effects of long-term air pollution in the general population [180]. A random sample of Swiss adults was selected from eight communities with different grades of urbanization. The first follow-up examination (SAPALDIA 2) took place from 2001 to 2003 [181] and the third survey was completed in 2011. SAPALDIA boasts of a very motivated study sample as more than 60% of the originally 9651 participants were still part of the third survey 20 years later (4934 participated at the respective study centre, and an additional 1205 participated at home over the telephone or by filling in questionnaires).The main health-related focus over the past 20 years was set on understanding the development and progression of respiratory diseases and CVDs.

3.1.2   Data

The different types of data contributing to the SAPALDIA research potential are depicted in Figure 3. Due to the still on-going collection and cleaning process of SAPALDIA 3 data at the time of carrying out this PhD work, all the projects of the current thesis did not involve any data from the third survey.
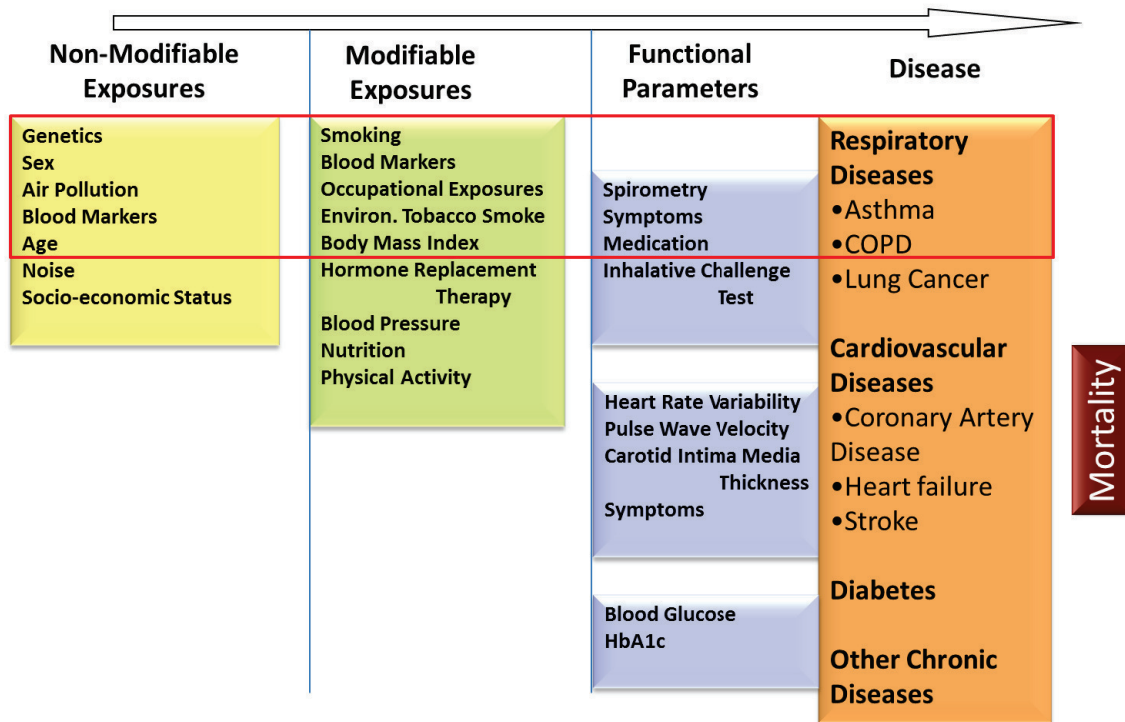
Non-Modifiable Exposures	Modifiable Exposures	Functional Parameters	Disease
Genetics **Sex** **Air Pollution** **Blood Markers** **Age**	**Smoking** **Blood Markers** **Occupational Exposures** **Environ. Tobacco Smoke** **Body Mass Index**	**Spirometry** **Symptoms** **Medication**	**Respiratory Diseases** •Asthma •COPD
Noise Socio-economic Status	Hormone Replacement Therapy Blood Pressure Nutrition Physical Activity	**Inhalative Challenge Test**	•Lung Cancer
		Heart Rate Variability **Pulse Wave Velocity** **Carotid Intima Media Thickness** **Symptoms**	**Cardiovascular Diseases** •Coronary Artery Disease •Heart failure •Stroke
		Blood Glucose **HbA1c**	**Diabetes** **Other Chronic Diseases**

Mortality

Figure 3. The SAPALDIA research potential. The most relevant data for this PhD work are depicted in the red rectangle.

Computer-assisted interviews were performed at the first two examinations comprising questions about the presence of asthma and respiratory symptoms, detailed personal smoking history, exposure to ETS at home or at the workplace, occupational exposure to vapours, gas, dusts and fumes, anthropometric factors and other data less relevant for this PhD work. Asthma-related questions included a self-report of having ever suffered from asthma and statements about a diagnosis by a physician and the recent use of medication. Lung function testing was performed with identical spirometers at both examinations [182]. Resource constraints only allowed spirometry without the application of bronchodilation. Expiratory flows were taken from the flow-volume curve with the highest sum of FVC and FEV_1. Blood markers (e.g. AAT and CRP) and genetic data were only measured in follow-up surveys. Longitudinal blood-derived data could hence not be used in the projects of this thesis. The genotyping of SNPs in candidate genes was performed by fluorescent 5' nuclease real-time PCR in dedicated laboratories in Zürich, Pavia and Innsbruck. Roughly 6000 persons agreed to genetic testing. In the framework of the GABRIEL study, about a quarter of them (i.e. all self-reported asthmatics as well as a random sample of non-asthmatics) were genome-wide assessed on the Illumina Human610-Quad platform [79].

More than half a million genotyped SNPs passed the quality control measures, and another 2 million autosomal SNPs could be inferred by HapMap-based imputation. 1000 genomes-based imputation was only partly available at the time of carrying out the analyses. PM_{10} exposure, a health-relevant air pollution measure consisting of particles which deposit in the airways and the lung, was predicted based on a Gaussian dispersion model for the years 1990 and 2000 with a 200 m grid resolution throughout Switzerland [183]. Individual exposures were derived by mapping the geo-referenced residential address of each study participant to the corresponding model grid cell. Annual exposure values for other years were derived by inter- and extrapolating the estimates from 1990 and 2000 based on trends from fixed monitoring stations.

Taking together, SAPALDIA is a well-representative sample of the adult Swiss population and comprises little population heterogeneity due to language and duration of residence requirements in the sampling process. Some general methodological restrictions important for all the projects shall be mentioned here. Asthma status was not physiologically defined and was therefore subject to misclassification. COPD prevalence and incidence in a sample of middle-aged adults from the general population is moderate, which results in little statistical power when considering these outcomes. Moreover, the non-availability of post-bronchodilator spirometry did not allow us to properly define COPD. In fact, more than 20% of (GOLD-defined) obstructed SAPALDIA baseline participants lost obstruction during follow-up, which points to measurement error or misclassification [184]. The use of longitudinal lung function was therefore superior in terms of statistical power and validity. The possibly distorting influence of asthmatics was generally taken into account in sensitivity analyses. Follow-up participants were on average healthier at baseline than those lost to follow-up. Inverse probability weighting was therefore regularly applied in sensitivity analyses to take a possible participation bias into consideration. A large number of potential confounders were included in the regression analyses of all the projects, but on two factors with conceivable relevance to pulmonary health (namely physical activity and dietary habits), we had not collected longitudinal data.

4 Results: Cell Cycle Genes in the Obesity-Asthma Association

4.1 Paper 1: The Association of a Variant in the Cell Cycle Control Gene *CCND1* and Obesity on the Development of Asthma in the Swiss SAPALDIA Study.

This paper was published:

Thun GA, Imboden M, Berger W, Rochat T, Probst-Hensch NM. Journal of Asthma 2013; 50(2):147-154.

informa
healthcare

GENETICS

The Association of a Variant in the Cell Cycle Control Gene *CCND1* and Obesity on the Development of Asthma in the Swiss SAPALDIA Study

GIAN ANDRI THUN, M.S.,[1,2] MEDEA IMBODEN, PH.D.,[1,2] WOLFGANG BERGER, PH.D.,[3] THIERRY ROCHAT, M.D.,[4] AND NICOLE M. PROBST-HENSCH, PH.D.,[1,2,*]

[1]*Swiss Tropical and Public Health Institute, Basel, Switzerland.*
[2]*University of Basel, Basel, Switzerland.*
[3]*Institute of Medical Molecular Genetics, University of Zurich, Zurich, Switzerland.*
[4]*Division of Pulmonary Medicine, University Hospital of Geneva, Geneva, Switzerland.*

Objective. The molecular mechanisms underlying the association between obesity (BMI ≥ 30 kg/m^2) and asthma are poorly understood. Since shifts in the fate of bronchial cells due to low-grade systemic inflammation may provide a possible explanation, we investigated whether two of the best documented functional variants in cell cycle control genes modify the obesity–asthma association. *Methods.* We genotyped 5930 SAPALDIA cohort participants for the single-nucleotide polymorphisms (SNPs) rs9344 in the cyclin D1 gene (*CCND1*) and rs1042522 in the gene encoding tumor protein 53 (*TP53*). We assessed the independent association of these SNPs and obesity with asthma prevalence and incidence. *Results.* The *CCND1* SNP modified the association between obesity and asthma prevalence ($p_{interaction} = 0.03$). The odds ratios (ORs) and 95% confidence intervals (CIs) for reporting a physician diagnosis of asthma at baseline, comparing obese with non-obese participants, were 1.09 (0.51–2.33), 1.64 (0.94–2.88), and 3.51 (1.63–7.53) for GG, GA, and AA genotypes, respectively. We found comparable genotype differences for incident asthma within the 11 years of follow-up. As for the *TP53* SNP, the interactions with obesity status with respect to asthma were not statistically significant. *Conclusions.* Our results suggest that obesity may contribute to asthma and associated tissue remodeling by modifying the processes related to the *CCND1* gene activity.

Keywords   candidate gene association study, cell proliferation genes, gene–lifestyle interaction, overweight, population-based cohort, single-nucleotide polymorphism

INTRODUCTION

In many parts of the world, obesity prevalence is still on the rise and represents a major public health concern. Obesity increases the risk for later asthma (1), but the underlying etiologic mechanisms for this epidemiologic association remain poorly understood. Explanations based solely on mechanistic effects of excessive body weight seem insufficient, as other aspects of the metabolic syndrome are also associated with impaired lung function (2). Obesity-related inflammatory markers or adipose derived hormones may influence bronchial hyperresponsiveness (BHR) as well as airway inflammation and associated tissue remodeling, as demonstrated in mouse models (3,4). But in human asthma studies, the data in favor of an effect of these circulating molecules on asthma risk remain inconclusive (5,6).

Investigating if gene variants of relevance to inflammatory pathways modify the obesity–asthma association may improve understanding of underlying etiologic mechanisms involved. Currently, there are very few studies available in adults which examine gene–obesity interactions with respect to asthma (7). Candidate genes not yet looked at include those essential to cell cycle control, as airway inflammation and remodeling in asthmatics are associated with altered apoptosis and proliferation of different bronchial cells types such as airway smooth muscle (ASM) and epithelial cells, fibroblasts, as well as cells of the immune system (8–11).

We therefore selected a functionally well-described polymorphism in each of two genes crucial for cell division in order to assess gene–obesity interactions in asthma in a highly focused manner. The cyclin D1 gene (*CCND1*) promotes cell proliferation through cell cycle G1–S phase transition, while the tumor protein 53 gene (*TP53*) is a tumor suppressor gene with a pivotal role in apoptosis. The expression of both genes is commonly altered in numerous cancer types. The single-nucleotide polymorphisms (SNPs) rs9344 [P242P] in *CCND1* and rs1042522 [R72P] in *TP53* are functionally very well characterized and have been associated with various cancer types, including in large meta-analyses (12–14). We previously demonstrated furthermore that these two polymorphisms modified the effect of particulate matter (PM$_{10}$) air pollution on lung function decline in the SAPALDIA cohort (15). In this same cohort representing the general population from eight Swiss communities, we now investigated whether these two SNPs also affect asthma risk and its relation to obesity. We argue that modification of the asthma–obesity association by variants in cell cycle control genes adds to the evidence for a causal obesity effect, mediated in part by inflammatory pathways.

*Corresponding author: Nicole Probst-Hensch, Ph.D., Head Unit Chronic Disease Epidemiology, Swiss Tropical and Public Health Institute, Socinstrasse 57, PO Box 4002 Basel, Switzerland; Tel: +41 61 284 83 78 (direct); +41 61 284 83 88 (admin); Fax: +41 61 284 81 05; E-mail: Nicole.Probst@unibas.ch

RIGHTS LINK

METHODS

Study Population

The methods of the SAPALDIA study have been described in detail elsewhere (16,17). In short, the study population consists of a random population sample of white adults aged 18–60 years from eight areas of Switzerland. Totally, 9651 persons participated in the first assessment in 1991, and at the follow-up examination in 2002, 8047 were reassessed with a least a short screening questionnaire. Blood donation and consent to genetic analyses at follow-up was given by 6058 participants of whom 6040 led to successful genotype determination. We could furthermore not include 110 subjects who either had missing information on asthma, smoking status or body mass index (BMI), or reported an asthma diagnosis which had not been confirmed by a physician. Taken together, 3721 individuals of the original study sample could not be considered for the present analysis (hereinafter called non-participants). The analyzed study sample consists hence of a total of 5930 participants. Written consent was obtained from all study participants separately for each assessment procedure. Approval of the study was given by the Swiss Academy of Medical Sciences and the regional ethics committees.

Definition of Smoking Status, Asthma Status, and Obesity

At both surveys, participants underwent a detailed, computer-assisted interview comprising questions about smoking behavior, exposure to environmental tobacco smoke, workplace exposure to dust and fumes, co-morbidities including asthma, medication use, and socio-economic factors. Never smokers were defined as persons who at the time of the interview had smoked less than 20 packs of cigarettes or 360 g of tobacco during their lifetime. Former smokers reported quitting at least one month before the interview. Asthmatics were defined in two ways. First, physician-diagnosed asthma was defined by affirmative answers to "Have you ever had asthma?" and "Was this confirmed by a doctor?" Second, current asthmatics, presenting a subset of those who were diagnosed by a physician, were classified by an affirmative answer to at least one of the following two questions: "Have you had an attack of asthma in the last 12 months?" and "Are you currently taking any medicines including inhalers, aerosols or tablets for asthma?" Incident asthma was defined as a new report of asthma among subjects without asthma at baseline. BMI was divided into four categories: low BMI (BMI < 20 kg/m^2), normal weight (20 kg/m$^2 \leq$ BMI < 25 kg/m^2), overweight (25 kg/m$^2 \leq$ BMI < 30 kg/m^2), and obese (BMI $\geq$ 30 kg/m^2).

Genotyping

DNA extraction from EDTA-buffered whole blood has been previously described (17). Originally, in each of three candidate genes pivotal for cell cycle control (*CCND1*, *TP53,* and *CDKN1A*, the gene encoding cyclin-dependent kinase inhibitor 1A, also called p21), one promising candidate SNP was selected according to its frequency and its reported functionality in cell division and apoptosis (rs9344 in *CCND1*, rs1042522 in *TP53*, and rs1801270 in *CDKN1A*). Genotyping was carried out using fluorescent 5-nuclease real-time PCR methodology (TaqMan, Applera Europe, Rotkreuz, Switzerland) and ABI Prism 7900 sequence detection system (ABI, Rotkreuz, Switzerland). The SNP-specific primers and LNA® dual-labeled fluorogenic probes were designed by Sigma Proligo (Evry, France). A 10% random sample of all DNA samples was re-genotyped, and all genotypes were confirmed. The genotype call rate was >99% for all the three SNPs. Hardy–Weinberg equilibrium was preserved apart from rs1801270 ($p = 0.04$). Since this latter SNP also showed a low minor allele frequency (MAF = 7%), statistical power to investigate effects on asthma and obesity was insufficient, and we excluded this SNP from the analyses of this study.

Statistical Analysis

Main effects of alleles and BMI on physician-diagnosed and current asthma at baseline were assessed using multivariate unconditional logistic regression. *CCND1* and *TP53* genotypes were generally included in a co-dominant fashion using the homozygous genotype of the more frequent allele as the reference category. Due to the limited number of *TP53* CC carriers, this category was combined with the group of heterozygous GC carriers for the analyses of incident asthma, assuming a dominant genetic model. All regression models included study area, sex, age at baseline, and smoking status (never vs. former vs. current) at baseline as covariates. These factors were chosen a priori as they may confound the obesity–asthma association. Models investigating the longitudinal effects of genotypes and change in BMI on incident asthma were additionally adjusted for smoking status at follow-up. Interaction between obesity and genotypes was tested by integrating multiplicative terms in the regression models. Significance levels for two-sided tests of main effects were chosen at $\alpha = 0.05$, and at $\alpha = 0.10$ for tests of effect modification. All statistical analyses were performed using STATA version 10.1 (StataCorp, College Station, TX, USA).

RESULTS

A detailed characterization of the study population can be found in Table 1. Study participants were about half a year older on average and had a slightly lower BMI than non-participants. Totally, 6.2% declared to have ever received an asthma diagnosis by a physician, whereas 2.6% suffered from current asthma. These percentages were higher in non-participants. Never smokers were more likely to be included in this study, confirming that a sample with higher health awareness was attracted to the follow-up examination.

The cross-sectional associations of the genetic polymorphisms in *CCND1* and *TP53* with the prevalence of asthma at baseline are presented in Table 2. The prevalence

TABLE 1.—Baseline characteristics of participants and non-participants in this study.

	Participants, $N = 5930$			Non-participants				
	Mean	SD	%	Mean	SD	%	N (with info)	*p*-Value
Female			50.2			51.8	3721	0.16
Age (years)	41.2	11.4		40.7	12.0		3721	0.06
BMI (kg/m^2)	23.8	3.6		24.2	4.2		3622	**<0.01**
BMI $\geq$ 30 kg/m^2			5.8			9.0	3622	**<0.01**
Physician-diagnosed asthma			6.2			7.8	3624	**<0.01**
Current asthma			2.6			3.4	3616	**0.02**
Smoking-status: never			47.3			38.5	3706	
Smoking-status: former			23.1			21.7	3706	**<0.01**
Smoking-status: current			29.6			39.8	3706	

Notes: Participants include all SAPALDIA follow-up subjects with valid information on sex, age, recruiting area, BMI, smoking status, and who could be successfully genotyped. Answers to questions about self- and physician-diagnosed asthma, as well as recent asthma attacks and asthma medication had to be complete and unambiguous. All SAPALDIA subjects from the original study sample ($N = 9651$) who could not be considered in the present analysis were combined and referred to as non-participants. Equal proportions were tested by chi-square test, equal distributions in age and BMI by Mann–Whitney test. SD, standard deviation; BMI, body mass index.

TABLE 2.—Associations of genetic polymorphisms with asthma prevalence at baseline.

Genotype	N	Prevalence (%)	OR (95% CI)	*p*-Value
Physician-diagnosed asthma				
CCND1, rs9344				
GG	1648	7.1	1	
GA	2942	6.1	0.85 (0.66–1.08)	0.17
AA	1340	5.3	0.73 (0.54–0.99)	**0.05**
TP53, rs1042522				
GG	3257	6.0	1	
GC	2291	6.5	1.10 (0.88–1.36)	0.42
CC	382	5.5	0.89 (0.56–1.42)	0.62
Current asthma				
CCND1, rs9344				
GG	1648	2.5	1	
GA	2942	2.6	1.05 (0.71–1.53)	0.82
AA	1340	2.8	1.09 (0.69–1.70)	0.72
TP53, rs1042522				
GG	3257	2.5	1	
GC	2291	3.0	1.21 (0.87–1.68)	0.26
CC	382	2.1	0.82 (0.39–1.70)	0.59

Notes: ORs were based on regression models adjusted for age, sex, study area, and smoking status. OR, odds ratio; CI, confidence interval.

of physician-diagnosed asthma was lower among *CCND1* rs9344 AA genotypes compared to homozygous GG individuals ($p = 0.05$), but this difference could not be detected for current asthma. The polymorphism in *TP53* was not associated with either asthma phenotype.

By dividing the individuals into four BMI categories, we confirmed the well-established association between asthma and obesity (Table 3). Compared to those with BMI between 20 and 25, obese individuals had a 1.81 times higher odds of having an asthma diagnosis by a physician ($p = 0.003$) and a 2.69 times higher odds of suffering from current asthma ($p < .001$). Interestingly, the cross-sectional association between asthma and obesity depended on the genotype of the *CCND1* SNP. The association of obesity with the risk of asthma was restricted to the AA genotype as graphically displayed in Figure 1 for physician-diagnosed asthma. Since low BMI and overweight did not essentially alter the odds of being an asthmatic of normal weight people, we classified all non-obese

subjects (BMI < 30 kg/m^2) into one category and compared them with obese subjects. Overall, obese subjects were 1.78 times (95% confidence interval (CI) 1.22–2.61) more likely to report an asthma diagnosis confirmed by a physician. For AA carriers of rs9344 in *CCND1*, this odds ratio (OR) increased to 3.51 (95% CI 1.63–7.53). The equivalent ORs for current asthma were even more pronounced (overall OR 2.65, 95% CI 1.60–4.40; and for AA carriers OR 7.17, 95% CI 2.89–17.83). Interaction terms between *CCND1* genotype and obesity were significant for both asthma definitions ($p = 0.03$ for physician-diagnosed asthma and $p = 0.1$ for current asthma). As for rs1042522 in *TP53*, obesity tended to be more strongly associated with the risk of asthma in homozygotes for the G allele than in heterozygotes, but these differences were based on very small sample sizes. When including BMI as a continuous variable, we observed similar patterns of associations.

Table 4 presents the longitudinal association of asthma incidence with the change in obesity status during follow-up. Obesity development was categorized into three distinct groups: participants who were never obese, participants who reached obesity at follow-up, and participants who were obese at both times of data collection. The latter group had the highest risk for developing asthma between baseline and follow-up (OR 2.05 for physician-diagnosed asthma and OR 2.58 for current asthma, compared to those who were never obese). Stratification by *CCND1* genotypes modified this risk. We observed the highest risk among AA carriers ($p < 0.001$ for both asthma definitions). Despite the wide confidence intervals, the *p*-values for interaction between *CCND1* genotype and obesity development in relation to asthma incidence were marginally statistically significant ($p = 0.09$ for physician-diagnosed asthma and $p = 0.03$ for current asthma). The *TP53* SNP did not modify the association between obesity development and asthma incidence.

Additional adjustment for atopy (positive skin prick test), smoking intensity, early childhood infection, family history of asthma, and exposure to gas, dust, or fumes did not substantially alter the results (data not shown).

TABLE 3.—Cross-sectional association of BMI with asthma risk at baseline, overall and stratified by *CCND1* and *TP53* genotypes.

	BMI	Physician-diagnosed asthma				Current asthma			
		Yes/no	OR	95% CI	*p*-Value	Yes/no	OR	95% CI	*p*-Value
All	<20	40/712	0.85	0.59–1.22	0.38	16/736	0.80	0.45 to 1.39	0.42
	20–25	195/3053	1	ref		81/3167	1	ref	
	25–30	98/1490	1.09	0.84–1.41	0.53	40/1548	1.11	0.75 to 1.66	0.60
	≥30	33/309	1.81	1.22–2.68	**0.003**	19/323	2.69	1.59 to 4.56	**<0.001**
	<30	333/5255	1	ref		137/5451	1	ref	
	≥30	33/309	1.78	1.22–2.61	**0.003**	19/323	2.65	1.60 to 4.40	**<0.001**
	Continuous	366/5564	1.05	1.02–1.08	**0.003**	156/5774	1.08	1.04 to 1.12	**<0.001**
***CCND1*, rs9344 GG**	<30	109/1433	1	ref		39/1503	1	ref	
	≥30	8/98	1.09	0.51–2.33	0.83	3/103	1.20	0.35 to 4.09	0.77
	Continuous	117/1531	0.98	0.93–1.04	0.55	42/1606	0.99	0.90 to 1.09	0.83
GA	<30	163/2607	1	ref		69/2701	1	ref	
	≥30	15/157	1.64	0.94–2.88	0.08	8/164	2.07	0.97 to 4.44	0.06
	Continuous	178/2764	1.06	1.02–1.10	**0.005**	77/2865	1.08	1.02 to 1.14	**0.006**
AA	<30	61/1215	1	ref		29/1247	1	ref	
	≥30	10/54	3.51	1.63–7.53	**0.001**	8/56	7.17	2.89 to 17.83	**<0.001**
	Continuous	71/1269	1.11	1.04–1.19	**0.003**	37/1303	1.17	1.07 to 1.28	**<0.001**
***TP53*, rs1042522 GG**	<30	172/2887	1	ref		68/2991	1	ref	
	≥30	23/175	2.41	1.50–3.87	**<0.001**	12/186	3.27	1.69 to 6.30	**<0.001**
	Continuous	195/3062	1.06	1.02–1.11	**0.002**	80/3177	1.08	1.02 to 1.14	**0.01**
GC	<30	142/2027	1	ref		63/2106	1	ref	
	≥30	8/114	1.04	0.49–2.19	0.92	5/117	1.59	0.62 to 4.11	0.34
	Continuous	150/2141	1.01	0.96–1.06	0.68	68/2223	1.06	0.99 to 1.13	0.08
CC	<30	19/341	1	ref		6/354	1	ref	
	≥30	2/20	2.45	0.46–13.03	0.29	2/20	8.29	1.05 to 65.36	**0.05**
	Continuous	21/361	1.16	1.01–1.32	**0.03**	8/374	1.34	1.06 to 1.69	**0.02**

Notes: Regression models were adjusted for age, sex, study area and smoking status. *p*-Values for interaction between *CCND1* genotype and obesity were **0.03** (physician-diagnosed asthma) and **0.01** (current asthma). *p*-Values for interaction between *CCND1* genotype and continuous BMI were **0.02** (physician-diagnosed asthma) and **0.03** (current asthma). *p*-Values for interaction between *TP53* genotype and obesity were 0.14 (physician-diagnosed asthma) and 0.64 (current asthma). *p*-Values for interaction between *TP53* genotype and continuous BMI were 0.86 (physician-diagnosed asthma) and 0.50 (current asthma). BMI, body mass index; OR, odds ratio; CI, confidence interval.
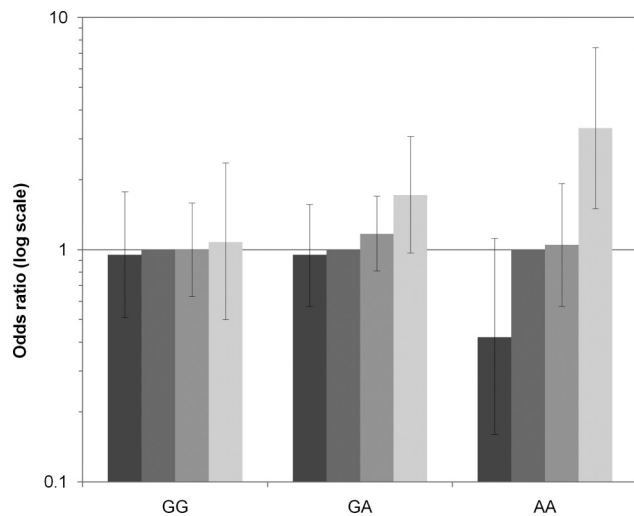
FIGURE 1.—Odds ratios (bars) and 95% confidence intervals (lines) for the cross-sectional association of body mass index (BMI) category and physician-diagnosed asthma stratified by *CCND1* genotype. BMI categories represent from darker to lighter color: <20 kg/m², 20–25 kg/m², 25–30 kg/m², and ≥30 kg/m². The reference group corresponded to individuals with normal weight (BMI 20–25 kg/m²).

DISCUSSION

We show for the first time that the cross-sectional and longitudinal association of obesity and asthma is modified by a SNP in a gene which exhibits a central role in cell cycle control. The associations appear to be stronger among subjects homozygous for the *CCND1* A allele in rs9344, the same genotype which in numerous studies including a large meta-analysis predisposed to various cancer subtypes across different ethnic populations (13). According to our study, these participants' lung tissue may be more prone to remodeling in response to obesity-related processes.

Asthma in Obese People

Our results confirm the well-established association of obesity with asthma risk. Current knowledge include that the obesity–asthma association decreases with progressive age (18) and may in part be mediated by low-grade systemic inflammation (19). Twin studies suggest that shared genetic pathways for asthma and obesity may partly explain the observed associations (20). Candidate gene studies and, more recently, genome-wide association studies have identified a large number of genes associated with either of the two conditions, but these genes do not seem to have a substantial impact on both phenotypes simultaneously (21). Candidate genes which seemed to be important for the asthma risk in obese people or for the higher BMI in asthmatics encompass *LEP*, *TNFA*, and *PRKCA*, regions encoding the adipose tissue derived

TABLE 4.—Longitudinal association of change in obesity status during follow-up with asthma incidence, overall and stratified by *CCND1* and *TP53* genotypes.

BMI (baseline/follow up)		Incidence of physician-diagnosed asthma (N = 174 cases)				Incidence of current asthma (N = 115 cases)			
		Yes/no	OR	95% CI	*p*-Value	Yes/no	OR	95% CI	*p*-Value
All	<30/<30	145/4481	1	ref		89/4735	1	ref	
	<30/≥30	15/551	0.91	0.53–1.57	0.73	15/588	1.47	0.84–2.57	0.18
	≥30/≥30	14/266	2.05	1.15–3.66	**0.02**	11/282	2.58	1.34–4.96	**0.005**
CCND1, rs9344									
GG	<30/<30	42/1231	1	ref		29/1309	1	ref	
	<30/≥30	2/140	0.46	0.11–1.94	0.29	1/158	0.30	0.04–2.28	0.24
	≥30/≥30	4/85	1.90	0.63–5.74	0.25	3/90	1.73	0.50–6.05	0.41
GA	<30/<30	80/2212	1	ref		45/2344	1	ref	
	<30/≥30	11/272	1.25	0.65–2.40	0.51	10/287	2.01	0.99–4.09	0.05
	≥30/≥30	5/138	1.18	0.46–3.00	0.73	4/146	1.72	0.60–4.94	0.32
AA	<30/<30	23/1038	1	ref		15/1082	1	ref	
	<30/≥30	2/139	0.68	0.16–3.01	0.61	4/143	2.24	0.71–7.11	0.17
	≥30/≥30	5/43	8.86	2.82–27.81	**<0.001**	4/46	11.85	3.32–42.24	**<0.001**
TP53, rs1042522									
GG	<30/<30	85/2443	1	ref		53/2576	1	ref	
	<30/≥30	10/318	0.94	0.48–1.85	0.73	9/341	1.27	0.62–2.63	0.62
	≥30/≥30	9/153	2.10	1.01–4.37	**0.02**	8/164	2.85	1.29–6.26	**0.004**
GC/CC	<30/<30	60/2038	1	ref		36/2159	1	ref	
	<30/≥30	5/233	0.81	0.32–2.07	0.66	6/247	1.68	0.69–4.12	0.25
	≥30/≥30	5/113	1.90	0.73–4.92	0.19	3/118	1.96	0.58–6.64	0.28

Notes: Included are participants without asthma at baseline (*N* = 5472 for physician-diagnosed asthma and *N* = 5720 for current asthma). Regression models were adjusted for age, sex, study area, and smoking status at baseline and follow-up. GC and CC genotypes of rs1042522 were taken together since the number of these genotypes was very low. *p*-Values for interaction between *CCND1* genotype and obesity development were **0.09** (incidence of physician-diagnosed asthma) and **0.03** (incidence of current asthma). *p*-Values for interaction between *TP53* genotype and obesity development were 0.79 (incidence of physician-diagnosed asthma) and 0.68 (incidence of current asthma). BMI, body mass index; OR, odds ratio; CI, confidence interval.

molecules leptin and tumor necrosis factor alpha (TNF*α*) as well as the ubiquitously expressed protein kinase C alpha, a factor associated with adipocyte differentiation and insulin signaling (7,22,23). Cell cycle control genes have so far not been investigated in this respect, but their importance in inflammatory processes confer them etiologic plausibility to potentially modify the obesity–asthma relationship. Interestingly, an in vitro study with asthma serum-sensitized human ASM cells recently found protein kinase C alpha to upregulate cyclin D1 expression (24).

Asthma: High Inflammatory Stress and Altered Cell Cycle Control

In the present study, the *CCND1* polymorphism rs9344 was associated with physician-diagnosed asthma at borderline significance level, but not with current asthma. This effect seems hence to originate from a history of asthma or a form of asthma which does not manifest itself in attacks on a regular basis. The inflammatory processes in the bronchial tissue of untreated asthma lead to structural changes in the airways. Gene array techniques comparing gene expression profiles of various cell types in bronchial tissue as well as in peripheral blood between asthmatics and healthy controls exhibit significant differences in numerous genes (25). For instance, atopic asthma has been associated with both reduced apoptosis of airway inflammatory cells as well as reduced net *TP53* activity and thus reduced apoptosis of peripheral blood mononuclear cells (8,26). ASM mass and cell proliferation is

increased in asthma (9), and this feature seems to involve *CCND1* the expression of which was elevated in asthma serum-sensitized human ASM cells (24). A significant suppression of bronchial epithelial cell proliferation associated with increased *CDKN1A* expression was observed in the asthmatic bronchial epithelium in humans (27). Inherited differences in the cell cycle response to inflammatory and other oxidative stressors could hence in part underlie asthma etiology. In agreement, asthma risk factors with oxidative properties like tobacco smoke or air pollutants were shown to alter cell proliferation in the airways (28) and to alter lung function in a manner dependent on cell cycle gene variants (15).

Obesity and Molecular Pathways of Potential Relevance to Asthma

Obesity shares with some of the inhaled asthma triggers the capacity to induce a status of low-grade systemic inflammation. Diseases associated with over-nutrition are characterized by alterations in circulating levels of inflammatory cytokines and adipose derived hormones (29). This so-called metabolic inflammation interferes with the regulation of intracellular molecular pathways that include cell cycle control mechanisms and could therefore also play a role in asthma etiology.

The PI3K/Akt signal pathway is activated by many of the factors which are altered in obesity and lead to increased cell survival (30). This pathway is also crucial in asthma pathophysiology. Namely, it has recently been

RIGHTSLINK()

shown that osteopontin, an extracellular matrix protein upregulated in the lungs of asthmatics, activates the PI3K/Akt pathway, and is associated with airway remodeling and disease severity in human asthma (31,32).

The transcription factor nuclear factor kappa B (NF-κB), one of the downstream targets of PI3K/Akt, interacts with several adipose-derived molecules. For instance, leptin stimulates and adiponectin reduces NF-κB signaling in endothelial cells (33,34), while the cytokine TNFα has been shown to interact with NF-κB signaling in ASM cells (35). We previously reported a complex joint effect of obesity with a *TNFA* polymorphism on asthma in two large cohort studies including SAPALDIA (7). NF-κB also exhibits a central role in the proliferating airway epithelium of asthmatics (36). Its interplay with different cell cycle control genes including *TP53* and *CCND1* is diverse (37,38). Compatible with this picture, leptin and adiponectin have also been found to be able to modulate expression levels of cyclin D1 or p53 in different types of cancer cells (39,40).

Both adiponectin and leptin are stimuli of the AMP-activated protein kinase (AMPK) (41). AMPK serves as an energy sensor and suppresses cell proliferation in non-malignant and tumor cells by interacting with the cell cycle machinery (42). As cell growth and proliferation are energy-intensive processes, AMPK may act as an energy checkpoint, permitting progression through the cell cycle only in the presence of sufficient energy reserves. Interestingly, metformin, an AMPK activator used in the treatment of obesity-related diabetes, was found to inhibit ASM cell proliferation (43).

Since obesity is also an established risk factor for colon cancer and many of the described pathways and mechanisms are also discussed in the etiology of that disease (30), results from colon cancer epidemiology are relevant to the interpretation of our findings, as they support the notion that cell cycle genes interact with inflammatory and other oxidative stressors including obesity. We reported an interacting effect on the risk of colorectal cancer for the *CCND1* SNP studied in this work with dietary antioxidants and proteins exhibiting antioxidative properties (44). Taken together, these findings corroborate the view that (a) obesity-mediated inflammation may directly affect cell cycle control, proliferation, and apoptosis in a variety of different tissues and that (b) the efficiency of these effects can depend on polymorphisms in cell cycle control master regulatory genes.

Strengths and Limitations of This Study

This study has several advantages. The study sample was selected to be representative of the adult population in the eight study areas. The study population is well characterized and comparatively large. Population stratification is at best of minor influence since participants had to be local residents for at least 3 years prior to the first survey, had to show a good command of one of the national languages, and there was no difference in genotype frequencies between Swiss and non-Swiss citizens ($p = 0.88$ for

rs9344 in *CCND1* and $p = 0.26$ for rs1042522 in *TP53*). Moreover, stratification of the associations by study center or language region did not materially alter the main findings. The assessment of asthma relied on internationally validated questions identical to those used in the European Respiratory Health Survey (45).

Nevertheless, some limitations apply. Like in many studies on asthma the definitions of asthma relied solely on the self-report of asthma diagnosis, attacks, and medications. Asthma status may therefore be subject to misclassification. Furthermore, there is indication that obesity-related asthma is a clearly distinct phenotype from general asthma and should be addressed differently. In fact, traditional inflammatory mechanisms in the airways have not been found to be relevant (6) and common asthma medication is far less helpful in obese subjects (46). Moreover, the usually observed asthma remission after weight loss (47) does not seem to support extensive airway remodeling. However, it has been recently demonstrated that airway remodeling may persist in asthmatics with complete asthma remission (48).

We cannot exclude participation bias. Smokers and patients with asthma and related phenotypes were less likely to participate at follow-up. Respiratory health of study participants is therefore slightly better than in non-participants. However, unless participation of asthmatics and non-asthmatics at baseline and follow-up was influenced by one of the genotypes investigated, substantial bias of the results due to non-participation is unlikely. A further limitation is the absence of adequate data on physical activity. This does not allow us to properly differentiate between modification of obesity and physical activity by the genetic polymorphisms investigated.

The study focused on selected SNPs in the genes of interest by giving priority to SNPs intensively studied in cancer. The *CCND1* rs9344 A allele leads to modulated splicing and consequently elevated production of the cyclin D1b isoform, a transcript with a higher cellular transformation potential (49,50). Even though this isoform is not solely dependent on rs9344, this polymorphism has been associated with different types of cancer (12). Apart from the high-penetrance mutations in *TP53* that lead to Li–Fraumeni syndrome, the non-synonymous SNP rs1042522 is arguably the most likely with functional relevance (14). The G allele is more powerful in inducing apoptosis (51), and an analysis summarizing several studies looking at different types of cancer could find a statistically marginally higher risk for carriers of the C allele (52). Further functional studies on the protein level are needed to clarify how the *CCND1* and *TP53* polymorphisms influence the cell division activity.

Finally, despite the large sample size of our study, the statistical power became limited in some of the stratified analyses, especially for the *TP53* genotypes and asthma incidence. Replication of the results in an independent cohort study is therefore essential, and with the help of pathway analysis involving many more SNPs and genes, the importance of cell cycle control activity in the asthma–obesity association could be more rigorously assessed.

CONCLUSION

Since obesity is a worldwide rapidly growing phenomenon, it is of high public health relevance to clarify causality and mechanisms of its association with asthma. Our results, if confirmed, suggest that obesity may contribute to asthma and associated tissue remodeling by modifying processes related to the *CCND1* gene activity.

ACKNOWLEDGMENTS

DECLARATION OF INTEREST

REFERENCES

1. Beuther DA, Sutherland ER. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. Am J Respir Crit Care Med 2007; 175:661–666.

2. Leone N, Courbon D, Thomas F, Bean K, Jego B, Leynaert B, Guize L, Zureik M. Lung function impairment and metabolic syndrome: the critical role of abdominal obesity. Am J Respir Crit Care Med 2009; 179:509–516.

3. Shore SA, Schwartzman IN, Mellema MS, Flynt L, Imrich A, Johnston RA. Effect of leptin on allergic airway responses in mice. J Allergy Clin Immunol 2005; 115:103–109.

4. Medoff BD, Okamoto Y, Leyton P, Weng M, Sandall BP, Raher MJ, Kihara S, Bloch KD, Libby P, Luster AD. Adiponectin deficiency increases allergic airway inflammation and pulmonary vascular remodeling. Am J Respir Cell Mol Biol 2009; 41:397–406.

5. Sood A, Ford ES, Camargo CAJr. Association between leptin and asthma in adults. Thorax 2006; 61:300–305.

6. Jartti T, Saarikoski L, Jartti L, Lisinen I, Jula A, Huupponen R, Viikari J, Raitakari OT. Obesity, adipokines and asthma. Allergy 2009; 64:770–777.

7. Castro-Giner F, Kogevinas M, Imboden M, de Cid R, Jarvis D, Machler M, Berger W, Burney P, Franklin KA, Gonzalez JR, Heinrich J, Janson C, Omenaas E, Pin I, Rochat T, Sunyer J, Wjst M, Anto JM, Estivill X, Probst-Hensch NM. Joint effect of obesity and TNFA variability on asthma: two international cohort studies. Eur Respir J 2009; 33:1003–1009.

8. Vignola AM, Chanez P, Chiappara G, Siena L, Merendino A, Reina C, Gagliardo R, Profita M, Bousquet J, Bonsignore G. Evaluation of apoptosis of eosinophils, macrophages, and T lymphocytes in mucosal biopsy specimens of patients with asthma and chronic bronchitis. J Allergy Clin Immunol 1999; 103:563–573.

9. Johnson PR, Roth M, Tamm M, Hughes M, Ge Q, King G, Burgess JK, Black JL. Airway smooth muscle cell proliferation is increased in asthma. Am J Respir Crit Care Med 2001; 164:474–477.

10. Kraft M, Lewis C, Pham D, Chu HW. IL-4, IL-13, and dexamethasone augment fibroblast proliferation in asthma. J Allergy Clin Immunol 2001; 107:602–606.

11. Cohen L, Xueping E, Tarsi J, Ramkumar T, Horiuchi TK, Cochran R, DeMartino S, Schechtman KB, Hussain I, Holtzman MJ, Castro M. Epithelial cell proliferation contributes to airway remodeling in severe asthma. Am J Respir Crit Care Med 2007; 176:138–145.

12. Knudsen KE, Diehl JA, Haiman CA, Knudsen ES. Cyclin D1: polymorphism, aberrant splicing and cancer risk. Oncogene 2006; 25:1620–1628.

13. Pabalan N, Bapat B, Sung L, Jarjanazi H, Francisco-Pabalan O, Ozcelik H. Cyclin D1 Pro241Pro (CCND1-G870A) polymorphism is associated with increased cancer risk in human populations: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2008; 17:2773–2781.

14. Whibley C, Pharoah PD, Hollstein M. p53 polymorphisms: cancer implications. Nat Rev Cancer 2009; 9:95–107.

15. Imboden M, Schwartz J, Schindler C, Curjuric I, Berger W, Liu SL, Russi EW, Ackermann-Liebrich U, Rochat T, Probst-Hensch NM. Decreased PM10 exposure attenuates age-related lung function decline: Genetic variants in p53, p21, and CCND1 modify this effect. Environ Health Perspect 2009; 117:1420–1427.

16. Martin BW, Ackermann-Liebrich U, Leuenberger P, Kunzli N, Stutz EZ, Keller R, Zellweger JP, Wuthrich B, Monn C, Blaser K, Bolognini G, Bongard JP, Brandli O, Braun P, Defila C, Domenighetti G, Grize L,

Karrer W, Keller-Wossidlo H, Medici TC, Peeters A, Perruchoud AP, Schindler C, Schoeni MH, Schwartz J, Solari G, Tschopp JM, Villiger B. SAPALDIA: Methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. Soz Praventivmed 1997; 42:67–84.

17. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, Stutz EZ, Bayer-Oglesby L, Baum F, Brandli O, Brutsche M, Downs SH, Keidel D, Gerbase MW, Imboden M, Keller R, Knopfli B, Kunzli N, Nicod L, Pons M, Staedele P, Tschopp JM, Zellweger JP, Leuenberger P. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991–2003: methods and characterization of participants. Soz Praventivmed 2005; 50:245–263.

18. Lang JE, Hossain J, Dixon AE, Shade D, Wise RA, Peters SP, Lima JJ. Does age impact the obese asthma phenotype? Longitudinal asthma control, airway function, and airflow perception among mild persistent asthmatics. Chest 2011; 140:1524–1533.

19. Shore SA. Obesity, airway hyperresponsiveness, and inflammation. J Appl Physiol 2010; 108:735–743.

20. Hallstrand TS, Fischer ME, Wurfel MM, Afari N, Buchwald D, Goldberg J. Genetic pleiotropy between asthma and obesity in a community-based sample of twins. J Allergy Clin Immunol 2005; 116:1235–1241.

21. Melen E, Himes BE, Brehm JM, Boutaoui N, Klanderman BJ, Sylvia JS, Lasky-Su J. Analyses of shared genetic factors between asthma and obesity in children. J Allergy Clin Immunol 2010; 126(631–637): e631-638.

22. Szczepankiewicz A, Breborowicz A, Sobkowiak P, Popiel A. Are genes associated with energy metabolism important in asthma and BMI?. J Asthma 2009; 46:53–58.

23. Murphy A, Tantisira KG, Soto-Quiros ME, Avila L, Klanderman BJ, Lake S, Weiss ST, Celedon JC. PRKCA: A positional candidate gene for body mass index and asthma. Am J Hum Genet 2009; 85:87–96.

24. Du CL, Xu YJ, Liu XS, Xie JG, Xie M, Zhang ZX, Zhang J, Qiao LF. Up-regulation of cyclin D1 expression in asthma serum-sensitized human airway smooth muscle promotes proliferation via protein kinase C alpha. Exp Lung Res 2010; 36:201–210.

25. Hansel NN, Diette GB. Gene expression profiling in human asthma. Proc Am Thorac Soc 2007; 4:32–36.

26. Brutsche MH, Brutsche IC, Wood P, Brass A, Morrison N, Rattay M, Mogulkoc N, Simler N, Craven M, Custovic A, Egan JJ, Woodcock A. Apoptosis signals in atopy and asthma measured with cDNA arrays. Clin Exp Immunol 2001; 123:181–187.

27. Puddicombe SM, Torres-Lozano C, Richter A, Bucchieri F, Lordan JL, Howarth PH, Vrugt B, Albers R, Djukanovic R, Holgate ST, Wilson SJ, Davies DE. Increased expression of p21(waf) cyclin-dependent kinase inhibitor in asthmatic bronchial epithelium. Am J Respir Cell Mol Biol 2003; 28:61–68.

28. Broekema M, ten Hacken NH, Volbeda F, Lodewijk ME, Hylkema MN, Postma DS, Timens W. Airway epithelial changes in smokers but not in ex-smokers with asthma. Am J Respir Crit Care Med 2009; 180:1170–1178.

29. Fantuzzi G. Adipose tissue, adipokines, and inflammation. J Allergy Clin Immunol 2005; 115:911–919; quiz 920.

30. Huang XF, Chen JZ. Obesity, the PI3K/Akt signal pathway and colon cancer. Obes Rev 2009; 10:610–616.

31. Simoes DC, Xanthou G, Petrochilou K, Panoutsakopoulou V, Roussos C, Gratziou C. Osteopontin deficiency protects against airway remodeling and hyperresponsiveness in chronic asthma. Am J Respir Crit Care Med 2009; 179:894–902.

32. Samitas K, Zervas E, Vittorakis S, Semitekolou M, Alissafi T, Bossios A, Gogos H, Economidou E, Lotvall J, Xanthou G, Panoutsakopoulou V, Gaga M. Osteopontin expression and relation to disease severity in human asthma. Eur Respir J 2011; 37:331–341.

33. Bouloumie A, Marumo T, Lafontan M, Busse R. Leptin induces oxidative stress in human endothelial cells. FASEB J 1999; 13:1231–1238.

34. Ouchi N, Kihara S, Arita Y, Okamoto Y, Maeda K, Kuriyama H, Hotta K, Nishida M, Takahashi M, Muraguchi M, Ohmoto Y, Nakamura T, Yamashita S, Funahashi T, Matsuzawa Y. Adiponectin an adipocyte-derived plasma protein, inhibits endothelial NF-kappaB signaling through a cAMP-dependent pathway. Circulation 2000; 102:1296–1301.

35. Amrani Y, Chen H, Panettieri RAJr. Activation of tumor necrosis factor receptor 1 in airway smooth muscle: a potential pathway that modulates bronchial hyper-responsiveness in asthma?. Respir Res 2000; 1:49–53.

36. Janssen-Heininger YM, Poynter ME, Aesif SW, Pantano C, Ather JL, Reynaert NL, Ckless K, Anathy V, van der Velden J, Irvin CG, van der Vliet A. Nuclear factor kappaB, airway epithelium, and asthma: avenues for redox control. Proc Am Thorac Soc 2009; 6:249–255.

37. Webster GA, Perkins ND. Transcriptional cross talk between NF-kappaB and p53. Mol Cell Biol 1999; 19:3485–3495.

38. Witzel II, Koh LF, Perkins ND. Regulation of cyclin D1 gene expression. Biochem Soc Trans 2010; 38:217–222.

39. Chen C, Chang YC, Liu CL, Chang KJ, Guo IC. Leptin-induced growth of human ZR-75-1 breast cancer cells is associated with up-regulation of cyclin D1 and c-Myc and down-regulation of tumor suppressor p53 and p21WAF1/CIP1. Breast Cancer Res Treat 2006; 98:121–132.

40. Mistry T, Digby JE, Desai KM, Randeva HS. Leptin and adiponectin interact in the regulation of prostate cancer cell growth via modulation of p53 and bcl-2 expression. BJU Int 2008; 101:1317–1322.

41. Luo Z, Saha AK, Xiang X, Ruderman NB. AMPK the metabolic syndrome and cancer. Trends Pharmacol Sci 2005; 26:69–76.

42. Motoshima H, Goldstein BJ, Igata M, Araki E. AMPK and cell proliferation – AMPK as a therapeutic target for atherosclerosis and cancer. J Physiol 2006; 574:63–71.

43. Ratnovsky A, Mellema M, An SS, Fredberg JJ, Shore SA. Airway smooth muscle proliferation and mechanics: Effects of AMP kinase agonists. Mol Cell Biomech 2007; 4:143–157.

44. Probst-Hensch NM, Sun CL, Van Den Berg D, Ceschi M, Koh WP, Yu MC. The effect of the cyclin D1 (CCND1) A870G polymorphism on colorectal cancer risk is modified by glutathione-S-transferase polymorphisms and isothiocyanate intake in the Singapore Chinese Health Study. Carcinogenesis 2006; 27:2475–2482.

45. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. Eur Respir J 1994; 7:954–960.

46. Sutherland ER, Goleva E, Strand M, Beuther DA, Leung DY. Body mass and glucocorticoid response in asthma. Am J Respir Crit Care Med 2008; 178:682–687.

47. Eneli IU, Skybo T, Camargo CAJr. Weight loss and asthma: a systematic review. Thorax 2008; 63:671–676.

48. Broekema M, Timens W, Vonk JM, Volbeda F, Lodewijk ME, Hylkema MN, Ten Hacken NH, Postma DS. Persisting remodeling and less airway wall eosinophil activation in complete remission of asthma. Am J Respir Crit Care Med 2011; 183:310–316.

49. Betticher DC, Thatcher N, Altermatt HJ, Hoban P, Ryder WD, Heighway J. Alternate splicing produces a novel cyclin D1 transcript. Oncogene 1995; 11:1005–1011.

50. Solomon DA, Wang Y, Fox SR, Lambeck TC, Giesting S, Lan Z, Senderowicz AM, Conti CJ, Knudsen ES. Cyclin D1 splice variants. Differential effects on localization, RB phosphorylation, and cellular transformation. J Biol Chem 2003; 278:30339–30347.

51. Dumont P, Leu JI, Della Pietra AC3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. Nat Genet 2003; 33:357–365.

52. van Heemst D, Mooijaart SP, Beekman M, Schreuder J, de Craen AJ, Brandt BW, Slagboom PE, Westendorp RG. Variation in the human TP53 gene affects old age survival and cancer mortality. Exp Gerontol 2005; 40:11–15.

5 Results: Genetic Determinants of AAT Serum Level and their Interplay with the Environment on Determining Lung Function

5.1 Paper 2: Serum Levels and Genotype Distribution of Alpha1-Antitrypsin in the General Population.

This paper was published:

ORIGINAL ARTICLE

Serum levels and genotype distribution of α_1-antitrypsin in the general population

Ilaria Ferrarotti,[1] Gian Andri Thun,[2,3] Michele Zorzetto,[1] Stefania Ottaviani,[1] Medea Imboden,[2,3] Christian Schindler,[2,3] Arnold von Eckardstein,[4] Lucia Rohrer,[4] Thierry Rochat,[5] Erich W Russi,[6] Nicole M Probst-Hensch,[2,3] Maurizio Luisetti[1]

► Additional materials are published online only. To view these files please visit the journal online (http://thorax.bmj.com/content/67/8.toc).

[1]Center for Diagnosis of Inherited Alpha1-antitrypsin Deficiency, Department of Molecular Medicine, Section of Pneumology, IRCCS San Matteo Hospital Foundation, University of Pavia, Pavia, Italy
[2]Unit Chronic Disease Epidemiology, Swiss Tropical and Public Health Institute, Basel, Switzerland
[3]University of Basel, Basel, Switzerland
[4]Institute of Clinical Chemistry, University Hospital of Zurich, Zurich, Switzerland
[5]Division of Pulmonary Medicine, University Hospitals of Geneva, Geneva, Switzerland
[6]Division of Pulmonary Medicine, University Hospital of Zürich, Zürich, Switzerland

Correspondence to
Dr Maurizio Luisetti, IRCCS San Matteo Hospital Foundation, University of Pavia, Piazza Golgi 1, 27100 Pavia, Italy; m.luisetti@smatteo.pv.it

ABSTRACT

Rationale α1-Antitrypsin (AAT) deficiency is one of the commonest rare respiratory disorders worldwide. Diagnosis, assessment of risk for developing chronic obstructive pulmonary disease (COPD), and management of replacement therapy require the availability of precise and updated ranges for protein serum levels.
Objective This paper aims to provide ranges of serum AAT according to the main genotype classes in the general population.
Methods The authors correlated mean AAT serum levels with the main *SERPINA1* variants (M1Ala/M1Val (rs6647), M3 (rs1303), M2/M4 (rs709932), S (rs17580) and Z (rs28929474)) in 6057 individuals enrolled in the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) cohort.
Results The following ranges (5th—95th percentile) of AAT were found in the serum (g/litre): 1.050—1.640 for PI*MM, 0.880—1.369 for PI*MS, 0.730—1.060 for PI*SS, 0.660—0.997 for PI*MZ and 0.490—0.660 for PI*SZ. There was very little overlap in AAT serum levels between genotype classes generally not believed to confer an enhanced health risk (MM and MS) and those associated with an intermediate AAT deficiency and a potentially mildly enhanced health risk (SS, MZ).
Conclusion This work resulted in three important findings: technically updated and narrower serum ranges for AAT according to PI genotype; a suggestion for a population-based 'protective threshold' of AAT serum level, used in decision-making for replacement therapy; and more precise ranges framing the intermediate AAT deficiency area, a potential target for future primary prevention.

INTRODUCTION

One of the few unambiguously ascertained individual risk factors for chronic obstructive pulmonary disease (COPD) is the serum level of α1-antitrypsin (AAT), which in turn is strongly determined by the AAT genotype variant system, classically named PI type. A large body of evidence suggests that the degree of risk for COPD is inversely related to the serum AAT level according to the hierarchy PI NullNull > PI ZZ > PI SZ > PI MZ.[1–4] Therefore, accuracy in AAT serum level determination is a relevant factor in COPD risk assessment. Above the area termed 'severe' AAT deficiency (AATD), bounded by the AAT protective threshold level of 11 μM[5] and at high risk for developing COPD, lies the area of 'intermediate' AATD, whose threshold has not been determined but is currently used as a proxy for the PI*MZ genotype.

Key messages

What is the key question?
► What are the ranges of serum α1-antitrypsin (AAT) level in the general population?

What is the bottom line?
► State-of-the-art methodologies allowed identification of AAT ranges according to the major genotypes, narrower than those previously available. Moreover, the authors defined the intermediate AAT deficiency area (0.92—0.49 g/litre) of particular interest being a possible target for future interventional options and clarified the longstanding controversy in the conversion from μM to g/litre of the 'protective threshold' of AAT serum level, used in decision-making for replacement therapy.

Why read on?
► It is important to clearly identify the protective threshold for AAT deficiency and, in turn, the serum level of AAT characterising patients with severe AAT deficiency. It is also important to correctly diagnose patients with intermediate AAT deficiency (mostly with the PI*MZ genotype).

Notably, the currently used standard reference values for AAT in serum[6] show a broad and overlapping range of values for the PI MM, PI MZ, PI MS and PI SS classes and do not represent data from the general population. In the absence of such data, only AAT serum values below 11 μM are of use for the assessment of severe AATD and for COPD risk prediction, whereas meaningful reference values to classify intermediate AATD associated with different AATD genotypes are lacking. Careful evaluation of serum AAT concentration is the initial diagnostic test in patients with suspected AATD.[7] This measurement can be routinely performed in any clinical chemistry laboratory, and it is the determining factor that justifies further analysis such as genotyping and sequencing, which are performed in dedicated laboratories.[8] Thus, the need for updated reference intervals for AAT according to the different PI types is of clinical relevance. This is especially true for reference values related to PI MZ, which may also require clinical attention in the form of smoking counselling in light of evidence for an increased risk of developing airflow obstruction.[4] [9]

The aim of this paper was to correlate serum AAT levels with the main PI variants, using current standards of measurement and diagnosis, including the molecular characterisation of the *SERPINA1* gene encoding AAT.[8] To the best of our knowledge, this information has not been published for a large general population sample.

We took advantage of the first follow-up examination of the (Swiss Cohort Study on Air Pollution and Lung Diseases in Adults) SAPALDIA cohort, which included 8047 people randomly selected from eight population registries representing the three major Swiss language regions, including both urban and rural areas.[10] The SAPALDIA biobank, which includes blood and DNA samples for more than 6000 people, was used to perform a previous study on the *SERPINA1* molecular characterisation of 1399 samples displaying reduced serum AAT levels.[11] In this study all 6057 samples from subjects who gave consent for genetic analyses, including the 1399 mentioned above, were investigated for normal M and deficient Z and S *SERPINA1* variants. The data presented were used to define the prevalence of those variants and the levels of serum AAT according to the main genotypes in the general population.

MATERIALS AND METHODS
Subjects
The SAPALDIA cohort has been previously described.[10] At the baseline in 1991 the subjects, who were 18−60 years old and predominantly Caucasian of Swiss nationality, were randomly selected from eight population registries. The current cross-sectional investigation of serum AAT is restricted to follow-up data collected in 2002−2003 when the biobank was established and includes 6057 subjects who donated blood and consented to genetic analysis. The study was approved by the Central Ethics Committee of the Swiss Academy of Medical Science and Cantonal Ethics Committees for each of the eight examination areas.

Serum analysis
AAT (g/litre) and C-reactive protein (CRP, mg/litre) concentrations were determined by latex-enhanced immunoturbidimetric assay (COBAS Integra analyzer, Roche Diagnostics, Indianapolis, Indiana, USA), a robust assay with principles that are perfectly comparable to those of nephelometry.[12] The interassay coefficient of variation (CV) was 3.6−4.6%; lower detection thresholds for the AAT and CRP assays were 0.21 g/litre and 1 mg/litre, respectively, and reference values were 0.9−2.0 g/litre and <8 mg/litre, respectively. Each new batch of antiserum was compared with previous batches for value recovery and proportionality in actual assays. A clarified, delipidated, commercially available serum calibrant (Calibrator f.a.s. Proteins, Roche Diagnostics) was used during the study; the same calibration batch, buffers and other reagents were used throughout the entire study.

Single nucleotide polymorphism analysis
All subjects were typed for give SNPs: S (rs17580), Z (rs28929474), M1Ala/M1Val (rs6647), M3 (rs1303), M2/M4 (rs709932). Typing was performed by PCR with fluorescently labelled Taq-Man probes (Vic or Fam labels) on a LigthCycler480 (Roche Diagnostics). All single nucleotide polymorphisms (SNPs) were in Hardy−Weinberg equilibrium. Further details on SNP analysis are available in the online data supplement.

Detection of rare deficient variants
The presence of rare deficient mutations was determined by sequencing the coding region of the *SERPINA1* gene, as previously described, on selected samples as reported by Zorzetto and coworkers.[11]

Statistical analysis
AAT concentrations were normally distributed and analysis of variance (ANOVA) was applied to compare means in different subgroups. Reference values covered the range from the 5th to the 95th percentile of AAT serum values. Linear and quantile regression was used to calculate adjusted means and percentiles. Covariates in the regression models were selected according to a former publication[13] and they were all significantly associated with AAT concentrations. The receiver operating characteristic (ROC) curve was used to estimate the predictive accuracy of serum AAT, and maximation of the Youden index (ie, the sum of sensitivity and specificity minus 1) defined the optimal threshold for discrimination of genotype classes. Bootstrapping procedures were used to estimate the 95% CIs of the optimal thresholds. Statistical analysis was performed with MedCalc 9.4.2.0 (MedCalc Software, Mariakerke, Belgium), Stata V.10.1 IC and SAS V.9.2.

RESULTS
As a first step we identified the number of subjects belonging to different *SERPINA1* genotype classes and determined their frequency in the general population (table 1). The PI*MM genotype accounted for 5398 individuals (89.12% of the overall population), whereas PI*MS was the second genotype in order of frequency (7.48%), followed by PI*MZ (2.36%). Only one subject carrying the PI*ZZ genotype (0.02%) was identified. The two classes defined in table 1 as rare variants and novel variants, accounting for 42 subjects (0.69%), were very heterogeneous groups of variants and were therefore excluded from further analyses on the relationship between AAT serum levels and *SERPINA1* genotypes. Nevertheless, means and ranges of AAT serum levels did not notably change if rare and novel variants were not excluded (data not shown). The frequencies of S and Z alleles in the three main Swiss language groups (German, French, Italian) are shown in online table E1 and are further described in the online data supplement.

Table 1 Frequency of the *SERPINA1* genotype classes detected in 6057 subjects from the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA) cohort

Genotype	PI*MM†	PI*MS	PI*SS	PI*MZ	PI*SZ	PI*ZZ	Rare variants‡	Novel variants§
Number of subjects	5398	453	10	143	10	1	34	8
Frequency (%)	89.12	7.48	0.16	2.36	0.17	0.02	0.56	0.13

†The PI*MM genotype class encompasses different combinations of normal M variants (M1Ala/Val, M3, M2/M4). Details are reported in the online data supplement.
‡This class includes subjects heterozygous for rare deficient variants, such as I, P$_{lowell}$, M$_{malton}$, M$_{Wurzburg}$ etc. These data were previously analysed in more detail.[11]
§This class includes novel putative deficient variants detected during *SERPINA1* gene sequencing. These data were previously analysed in more detail.[11]

Table 2 Unadjusted and adjusted means and intervals (5th and 95th percentiles) for α1-antitrypsin (AAT) serum concentration in the six main *SERPINA1* genotype classes

SERPINA1 genotype	Unadjusted AAT serum concentrations, N=5981*				Adjusted AAT serum concentrations,† N=5768‡			
	N	Mean, SD (g/litre)	5th percentile	95th percentile	N	Mean (g/litre)	5th percentile	95th percentile
MM	5366	1.298, 0.18	1.05	1.64	5175	1.298	1.079	1.572
MS	451	1.085, 0.16	0.88	1.37	438	1.082	0.902	1.312
SS	10	0.849, 0.10	0.73	1.06	10	0.823	0.735	1.009
MZ	143	0.805, 0.11	0.66	1.00	136	0.811	0.672	1.011
SZ	10	0.555, 0.06	0.49	0.66	9	0.554	0.480	0.638
ZZ	1	0.320, 0.00						

*Subjects with rare variants (42) and samples with missing AAT levels (34) were excluded.
†Adjusted for age, sex, area, alcohol consumption (yes/no), systolic blood pressure, body mass index, smoking habit (never, former, current) and C-reactive protein levels.
‡Additionally excluded were subjects with missing covariate data (212) and ZZ genotype due to insufficient frequency (1).

Unadjusted and adjusted means and reference intervals for AAT serum concentration in the six main *SERPINA1* genotype classes are presented in table 2. Adjusting for age, sex, study area, alcohol intake (yes vs no), systolic blood pressure, body mass index (BMI), smoking status (never, former, current) and CRP levels did not essentially alter the results. The 5th–95th percentiles were subsequently compared with previously reported American Thoracic Society (ATS)/European Respiratory Society (ERS) reference values[6] (table 3). The AAT serum level ranges determined in our investigation are markedly more narrow than those previously reported. Data are also graphically reported in figure E1, in which AAT serum concentrations are mathematically converted to μM.

Since AAT is a acute phase protein, we recalculated the reference intervals of AAT serum concentrations in subgroups of subjects according to systemic inflammatory status. A CRP value of 8 mg/litre, which is the upper normal limit for this protein as suggested by the equipment used in the present study, was used as the cutoff to stratify subjects as being without (<8 mg/litre) or with (≥8 mg/litre) systemic inflammation (table E2). Comparison of AAT means between the two CRP strata revealed higher values in the systemic inflammation stratum for all genotype classes. This difference was statistically significant in the PI*MM subgroup (p<0.001).

As a next step we assessed the accuracy of predicting genotype classes which are not believed to represent a risk for developing emphysema (PI*MM and PI*MS) and those associated with intermediate AATD and arguably a slightly increased risk for developing emphysema (PI*SS and PI*MZ) from AAT concentrations using ROC statistics.[4 14] For this analysis we included the rare variant carriers of the respective groups to get

a representative sample for the general population. The findings were highly accurate for the area under the curve (AUC = 0.9907) (figure E2). The optimal threshold according to the Youden index provided a cutoff at 1.00 g/litre AAT level (95% CI 0.97 to 1.06), which presents a sensitivity of 95.8% and a specificity of 94.8%. For discrimination between PI*MM and any other genotype carrying at least one S or Z allele an optimal cutoff at 1.10 g/litre was determined (73.4% sensibility, 88.5% specificity). The impact of sex, smoking status and CRP levels on these genotype discriminations are described in the online data supplement (table E3).

Finally we analysed the influence of the different PI*M subtypes on AAT serum level. This result is reported in the online data supplement (table E4 and figure E3).

DISCUSSION

This study ideally represents the most valid setting to date to derive reference values for serum AAT by genotype group in the general population. We applied state-of-the-art technology for the assessment of serum AAT and *SERPINA1* genotypes in the Swiss population, which is a combination of three language groups that adequately represent the genetic structure of the European population.[15] To the best of our knowledge, only a few studies have been performed in the general population that measure circulating AAT protein[7] or *SERPINA1* gene variants[9] or both.[16 17] The most comparable study is the Copenhagen City Heart study, a longitudinal survey of 7963 subjects from Copenhagen who were genotyped for PI*Z and PI*S but in whom only a small sample of AAT concentrations in blood were measured (n=592).[9] In the study by Sveger,[16] blood from 200 000 infants was drawn for simultaneous AAT determination by semi-quantitative electroimmunoassay and Pi typing with isoelectric focusing. However, the analytical methods used in this study were out of date and therefore these data can no longer be used as a reference. The study by Silverman *et al*[17] applied an automated immunoassay to measure AAT in plasma samples from 20 000 blood donors from the St Louis area. Plasma samples that met criteria of <50% of plasma pool reactivity were examined by isoelectric focusing to determine PI type. The reported St Louis Z allele frequency was 0.0116, but no data on the reference values for the concentration of AAT in plasma were extrapolated. In summary, current AAT serum level–genotype relationships seem obsolete as updated diagnostic standards for AATD have never been applied to a general population sample.

Our work generated a number of outputs. First, a more precise allele and genotype frequency was identified for *SERPINA1* variants in Switzerland's general population. We calculated an updated Pi*S gene frequency of 0.0401 whereas that of PI*Z was 0.0130. Compared with previous estimates for this population,[18]

Table 3 Comparison between two published ranges (5th–95th percentiles) of α1-antitrypsin serum levels according to different phenotypes (PI),[1] one in μM and one in g/litre,[6] and the range, according to different genotypes (PI*), as deduced by our analysis[2]

Phenotype–genotype	Units	Reference ranges[1]	Present paper[2]
PI MM–PI*MM	μM	20–48	20.2–31.5
	g/l	1.50–3.50	1.05–1.64
PI SS–PI*SS	μM	15–33	14.0–20.4
	g/l	1.00–2.00	0.73–1.06
PI MZ–PI*MZ	μM	17–33	12.7–19.2
	g/l	0.90–2.10	0.66–1.00
PI SZ–PI*SZ	μM	8–16	9.4–12.7
	g/l	0.75–1.20	0.49–0.66

The g/litre values in our analysis were mathematically converted to μM, based on a molecular weight of 52 kDa. Note that the PI*MS data are not present because this genotype was not included in the original American Thoracic Society/European Respiratory Society guidelines.

PI*S frequency was similar whereas PI*Z frequency was slightly higher. Further discussion of this topic is provided in the online data supplement.

A second major outcome of the study was the analysis of AAT serum concentration and genotypes of the *SERPINA1* gene in a large cohort representative of the general population. This enabled the extrapolation of updated protein ranges according to the main normal and deficient genotype classes and evaluation of whether reference values need to be specific for sex, smoking status and CRP level.

To date, no epidemiological data on AAT serum concentration according to the *SERPINA1* genotype have been reported for the general population. Some studies identified mean values or reference intervals of AAT serum concentration but they were limited to target groups, mostly COPD cases or populations in which the frequency of AATD is low, such as Japanese[19] and Korean populations,[20] or in patients with a peculiar clinical phenotype, such as Peyronie's disease,[21] or characteristic cohorts, such as paediatric subjects[7] and pregnant women.[22] The use of advanced technologies makes this study innovative. These technologies include SNP detection for genotyping and the exclusion of potentially confounding genetic factors, that is, deficient variants other than S and Z alleles, from the final analysis. The presence of other variables that could affect the AAT concentration in serum were considered and adjusted for in an additional analysis. This resulted in much narrower serum AAT ranges than those presented in the ATS/ERS consensus document,[6] with a drastic reduction in the overlap among genotypes (table 3). Moreover, the mean values of AAT concentration according to *SERPINA1* genotypical classes (table 2), were lower than those reported so far,[9] likely due to a smaller upper dispersion of measurements. Although 70−80% of the variation in total AAT serum concentration is explained by the Pi type (after age and sex adjustment),[23] other factors can influence variation.[13] Since AAT is an acute-phase reactant, inflammatory status may increase the serum level of AAT.[11 13 24] As shown in table E2, when SAPALDIA subjects were stratified according to the presence or absence of an inflammatory condition, most *SERPINA1* genotypes showed even narrower ranges, particularly those with CRP <8 mg/litre. However, ranges for individuals with elevated CRP were generally higher. In the real world of routine AAT serum measurement, our data do not justify systematic measurement of the inflammatory status and the stratified ranges can be used only in reference laboratories.[8]

Other factors that impact AAT concentration variability include active smoking[25] and age.[21] An in-depth analysis of the same SAPALDIA population has shown that an inter-relationship among circulating AAT, smoke exposure, gender and systemic inflammatory status exists.[13] However, in this study with narrower diagnostic purposes, we demonstrated that inclusion of sex and current smoking status to predict intermediate deficiency genotypes was not necessary. We also addressed a further putative factor for AAT serum concentration variability, that is, the intrinsic effect of the different PI*M subtypes (figure E3). The effect of this variable seems to be negligible because only two PI*MZ haplotypes displayed significant changes in AAT serum concentration.

In this study, we also analysed the limits of the so-called 'protective threshold' and we tried to address the controversy and confusion about the expression of serum AAT concentration. The term 'protective threshold' derives from evidence that subjects with AATD and an AAT serum level above the threshold are at reduced risk of developing emphysema. This is not merely a theoretical cutoff because it is considered the decisional cutoff below which subjects with AATD are eligible for AAT replacement therapy.[6] Therefore, it is a very important concept in AATD patient management. An excellent discussion on this topic by Tonelli and Brantly has recently been published.[26] The concept was originally developed by Hutchinson *et al*[27] and Stockley,[28] based on evidence that subjects displaying the PI SZ phenotype had a reduced risk of developing emphysema compared with those displaying the PI ZZ phenotype[3] and are therefore seldom suitable for replacement therapy. The threshold was fixed at the serum AAT level corresponding to 0.8 g/litre, measured by radial immunodiffusion. A few years later, to resolve the lack of standardisation among laboratories that caused so much confusion in the definition of the AAT measurements, a highly purified AAT standard, expressed as μM, was introduced.[29] In the same report, the protective threshold using the highly purified AAT standard determined by nephelometry was fixed at 11 μM; that is, the 10th percentile of the AAT serum range for subjects with PI SZ, which is considered adequate to protect the lungs from proteolytic attack. Since then, in countries where the AAT concentration was expressed as g/litre, the 0.8 threshold was often considered equivalent to 11 μM. However, radial immunodiffusion cannot be considered equivalent to nephelometry because the former, obsolete method overestimates the real AAT concentration by about

Figure 1 Suggested areas corresponding to severe α1-antitrypsin (AAT) deficiency (below the protective threshold) and intermediate AAT deficiency (above the protective threshold and below the 10th percentile of the AAT range for subjects carrying the PI*MS genotype). Bars represent 5th/95th percentiles of AAT serum levels.
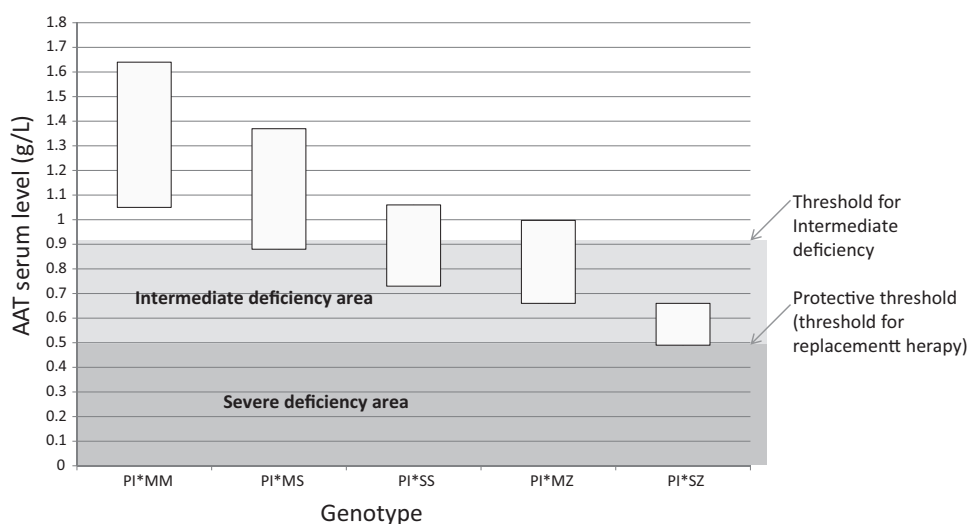
50%.[30] The value corresponding to the 10th percentile of AAT serum concentration for the PI*SZ group in SAPALDIA, which is suitable to derive threshold values in the general population, is 0.49 g/litre.

Intermediate deficiency is a term usually referred to as synonymous with the PI*MZ genotype, which may represent a slightly increased risk of developing COPD.[4] We believe that correct diagnosis of subjects carrying the PI*MZ genotype is a critical issue for a number of reasons. First, having been identified as a group at risk of developing COPD, they are subjects particularly suitable for an effective prevention and smoking cessation campaign, as suggested by an increased rate of attempting to quit smoking following genetic testing.[31] Second, correct diagnosis is mandatory for genetic counselling. Third, subjects with COPD carrying the PI*MZ genotype could be suitable for future, specific therapeutic interventions. Expressed as a range of serum levels, we propose that corresponding values stretch from the protective threshold (0.49 g/litre) to the 10th percentile of the AAT concentration range for subjects carrying the PI*MS genotype, who are believed not to be at risk of developing emphysema,[14] which would correspond to 0.92 g/litre in SAPALDIA. This area includes 87% of subjects carrying the PI*MZ genotype in our cohort. The reported thresholds and related areas are depicted in figure 1.

One of the aims of this paper was to provide a clear cutoff, below which suspicion of AATD is reasonable, and to resolve the controversy around this issue. The choice of the AAT cutoff, below which samples should be selected for PI pheno/geno-typing, has important financial and clinical implications. The cutoffs determined by individual laboratories currently range between 1.00 and 1.30 g/litre and they strongly depend on specific requirements. For example, the clinical importance of PI*MS detection is considered to be far less than PI*MZ detection due to the different risks for emphysema for the two genotypes.[4 14] Therefore, we reported two different cutoffs, one focused on avoiding the omission of deficient S or Z alleles (1.10 g/litre) and the second set to identify genotypes at a likely increased risk of emphysema (1.00 g/litre). We also considered the previously reported cutoff of 1.13 g/litre,[32] which is still useful since no Z alleles (and PI*SS) have been found in individuals with AAT blood levels above this level (100% sensitivity and 78.6% specificity for detecting AATD genotypes), while 31% of all assigned subjects with the PI*MS genotype show AAT blood levels higher than 1.13 g/litre (78.6% sensitivity and 82.6% specificity for detecting any deficient S or Z allele).

In conclusion, we provided values for serum AAT level according to the major genotype classes in the general population. In addition, these data have helped us to address controversies related to the different opinions in the definition of limits for the 'protective threshold' and to define a useful range for intermediate AATD. We believe that these findings will be helpful in the future for the investigation of AATD-related risk for COPD and for a more precise definition of when to implement AAT replacement therapy. Finally, the reported data show that the gene—environmental analysis is critical in the ongoing SAPALDIA longitudinal assessment of the impact of *SERPINA1* on pulmonary health.

REFERENCES

1. **Fregonese L**, Stolk J, Frants RR, et al. Alpha1-antitrypsin null mutations and severity of emphysema. *Respir Med* 2008;**102**:876—84.
2. **Ferrarotti I**, Baccheschi J, Zorzetto M, et al. Prevalence and phenotype of subjects carrying rare variants in the Italian registry for alpha1-antitrypsin deficiency. *J Med Genet* 2005;**42**:282—7.
3. **Turino GM,** Barker AF, Brantly ML, et al. Clinical features in individuals with PI*SZ phenotype of alpha1-antitrypsin deficiency. Alpha1-antitrypsin Deficiency Registry Study Group. *Am J Respir Crit Care Med* 1996;**154**:1718—25.
4. **Hersh CP,** Dahl M, Ly NP, et al. Chronic obstructive pulmonary disease in alpha 1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004;**59**:843—9.
5. **Wewers MD,** Casolaro MA, Sellers SE, et al. Replacement therapy for alpha1-antitrypsin deficiency associated with emphysema. *N Engl J Med* 1987;**316**:1055—62.
6. **American Thoracic Society, European Respiratory Society.** American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha1-antitrypsin deficiency. *Am J Respir Crit Care Med* 2003;**168**:818—900.
7. **Steiner SJ,** Gupta SK, Croffie JM, et al. Serum levels of alpha1-antitrypsin predict phenotypic expression of the alpha1-antitrypsin gene. *Dig Dis Sci* 2003;**48**:1793—7.
8. **Miravitlles M,** Herr C, Ferrarotti F, et al. Laboratory testing of individuals with severe alpha1-antitrypsin in three European centres. *Eur Respir J* 2010;**35**:960—8.

9. **Dahl M,** Tybjaerg-Hansen A, Lange P, et al. Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha1-antitrypsin MZ heterozygotes: a longitudinal study of the general population. Ann Intern Med 2002;**136**:270—9.

10. **Ackerman-Liebrich U,** Kuna-Dibbert B, Probst-Hensch NM, et al. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991—2003: methods and characterization of participants. Soz Praventivmed 2005;**50**:1—19.

11. **Zorzetto M,** Russi EW, Senn O, et al. SERPINA1 gene variants in subjects from the general population with reduced alpha1-antitrypsin level. Clin Chem 2008;**54**:1331—8.

12. **Ledue TB,** Collins MF. Development and validation of 14 human serum protein assays on the Roche cobas(®) c 501. J Clin Lab Anal 2011;**25**:52—60.

13. **Senn O,** Russi EW, Schindler C, et al. Circulating alpha 1-antitrypsin in the general population: determinants and association with lung function. Respir Res 2008;**9**:35.

14. **Dahl M,** Hersh CP, Ly NP, et al. The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J 2005;**26**:67—76.

15. **Lao O,** Lu TT, Nothnagel M, et al. Correlation between genetic and geographic structure in Europe. Curr Biol 2008;**18**:1241—8.

16. **Sveger T.** Liver disease in alpha1-antitrypsin deficiency detected by screening of 200,000 infants. N Engl J Med 1976;**294**:1316—21.

17. **Silverman EK,** Miletich JP, Pierce JA, et al. Alpha-1-Antitrypsin deficiency. High prevalence in the St. Louis area determined by direct population screening. Am Rev Respir Dis 1989;**140**:961—6.

18. **de Serres FJ.** Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. Chest 2002;**122**:1818—29.

19. **Kwok JS,** Lawton JW, Yew WW, et al. Protease inhibitor phenotypes and serum alpha 1-antitrypsin levels in patients with COPD: a study from Hong Kong. Respirology 2004;**9**:265—70.

20. **Kim CH,** Yim JJ, Yoo CG, et al. Alpha-antitrypsin genotypes in Korean patients with chronic obstructive pulmonary disease. Respirology 2005;**10**:223—8.

21. **Hauck EW,** Hauptmann A, Haag SM, et al. Alpha 1-antitrypsin levels and genetic variation of the alpha 1-antitrypsin gene in Peyronie's disease. Eur Urol 2004;**43**:623—8.

22. **Lisowska-Myjak B,** Sygitowicz G, Wolf B, et al. Serum alpha-1-antitrypsin concentration during normal and diabetic pregnancy. Eur J Obstet Gynecol Reprod Biol 2001;**99**:53—6.

23. **Silvermann EK,** Province MA, Campbell EJ, et al. Family study of a1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. Genet Epidemiol 1992;**9**:317—31.

24. **Ottaviani S,** Gorrini M, Scabini R, et al. C reactive protein and alpha 1-antitrypsin: relationship between levels and gene variants. Transl Res 2011;**157**:332—8.

25. **Ashley MJ,** Corey P, Chan-Yeung M. Smoking, dust exposure, and serum alpha 1-antitrypsin. Am Rev Respir Dis 1980;**121**:783—8.

26. **Tonelli AR,** Brantly ML. Augmentation therapy in alpha1-antitrypsin deficiency: advances and controversies. Ther Adv Respir Dis 2010;**4**:289—312.

27. **Hutchinson DC,** Tobin MJ, Cook PJ. Alpha 1 antitrypsin deficiency: clinical and physiological features in heterozygotes of Pi type SZ. A survey by the British Thoracic Association. Br J Dis Chest 1983;**77**:28—34.

28. **Stockley RA.** Proteolytic enzymes, their inhibitors and lung diseases. Clin Sci (Lond) 1983;**64**:119—26.

29. **Brantly ML,** Wittes JT, Vogelmeier CF, et al. Use of a highly purified a1-antitrypsin standard to establish ranges for the common normal and deficient a1-antitrypsin phenotypes. Chest 1991;**100**:703—7.

30. **Liappis N.** Study on the determination of IgA, IgG, IgM, α_1-antitrypsin, haptoglobin and transferrin with the kinetic nephelometric method. Comparison with the radial immunodiffusion. Klin Padiatr 1980;**192**:370—8.

31. **Carpenter MJ,** Strange C, Jones Y, et al. Does genetic testing result in behavioral health change? Changes in smoking behavior following testing for alpha1-antitrypsin deficiency. Ann Behav Med 2007;**33**:22—8.

32. **Gorrini M,** Ferrarotti I, Lupi A, et al. Validation of a rapid, simple method to measure alpha1-antitrypsin in human dried blood spots. Clin Chem 2006;**52**:899—901.

Thorax alerts

Sign up for our electronic table of contents alerts and you will never miss new issues of *Thorax* when published online. Stay ahead and up to date by visiting **thorax.bmj.com**.

ONLINE DATA SUPPLEMENT

Serum levels and genotype distribution of alpha$_1$-antitrypsin in the general population

Ilaria Ferrarotti[1] , Gian Andri Thun[2,3], Michele Zorzetto[1], Stefania Ottaviani [1], Medea Imboden [2,3],

Christian Schindler[2,3], Arnold von Eckardstein[4], Lucia Rohrer[4], Thierry Rochat[5] , Erich Russi[6],

Nicole M Probst-Hensch [2,3], Maurizio Luisetti [1]

SUPPLEMENTARY APPENDIX

Methods

SNP analysis

Results

Epidemiology of SERPINA1 Genotypes in Swiss language groups

ROC analysis for best cut-off calculation and accuracy of predicting genotype classes

Analysis of normal M variant subtypes

Discussion

Frequency of SERPINA1 Gene variants in Switzerland

AAT cut-off value for diagnostic purposes

Effect of SERPINA1 M variants on AAT serum concentration

Methods

SNP analysis

The PCR conditions were identical for all applications: 0.125 µl of 20X working stock of SNPGenotypingAssay, 2.5 µl LightCycler 480 Probes Master (Roche Diagnostics), and 20 ng DNA sample, in a total volume of 5 µl. PCR cycling conditions were also identical for all assays: initial denaturation step of 95°C for 10 minutes, followed by 40 cycles of denaturation at 95°C for 15 seconds, and annealing at 60°C for 1 minute. After the amplification, melting curves were generated by slowly (ramp rate 2.2°C/second) cooling the sample to 40°C. The sequences of primers and probes are available on request.

Results

Frequency of Z and S alleles in the Swiss language groups

As an aside, we analyzed the frequency of S and Z alleles in the three language groups: German (n=3,288), French (n=1,938) and Italian (n=831) (Table E1). Interestingly, while the PI*Z allele was homogeneously distributed among the three language groups, the frequency of the PI*S allele was significantly higher in the French subgroup (0.053) than in the German (0.036) and in the Italian (0.025) (p<0.001 for both comparisons).

ROC analysis for best cut-off calculation and accuracy of predicting genotype classes

We assessed the accuracy of predicting genotype classes which are not believed to represent a risk for developing emphysema (PI*MM and PI*MS) and those associated with intermediate AATD and arguably a slightly increased risk for developing emphysema, i.e. PI*SS and PI*MZ, from AAT concentrations using ROC statistics[E1,E2]. For this analysis we included the rare variant carriers of the respective groups in order to get a representative sample for the general population. The

findings were highly accurate for the area under the curve (AUC=0.9907) (Figure E2). The optimal threshold according to the Youden index provides the cut-off at 1.00g/L AAT level (95% CI 0.97 to 1.06), which represents a sensitivity of 95.8% and a specificity of 94.8%. As we had previously reported differences in circulating AAT concentrations by gender, smoking status and CRP levels[E3], we tried to assess whether prediction of PI*SS or MZ genotypes could be improved by considering the influence of these factors on AAT blood levels. We found a marginal enhancement of prediction quality when adjusting for sex and current smoking status, but not for CRP (AUC=0.9927, p=0.05). We subsequently recalculated values for normal (PI*MM and PI*MS) vs. intermediate deficiency genotype classes (PI*MZ and PI*SS) by gender and current smoking (Table E2). Despite slight differences in the stratified specific means, the impact of the genotype played a much bigger role than that of gender and current smoking.

For discrimination between PI*MM and any other genotype carrying at least one S or Z allele, an optimal cut-off at 1.10g/L was determined (73.4% sensibility, 88.5% specificity).

Analysis of normal M variant subtypes

The analysis of the three SNPs for normal variants (M1Ala/M1Val - rs6647; M3 - rs1303; M2/M4 - rs709932) by haplotype reconstruction revealed 14 normal genotypic classes in the PI*MM group, 5 classes in the PI*MS group and 5 classes in the PI*MZ group (Table E3). Among these, the most common were PI*M1(Val)M1(Val), PI*M1(Ala)M1(Val), and M1(Val)M2 (frequencies of 0.28, 0.21, and 0.15, respectively). The reference intervals (5th-95th percentiles) for AAT serum concentration have been calculated in each group (Figure E3). Comparison of means within each group revealed only a significant difference between PI*M1(Val)Z and PI*M1(Ala)Z (0.819 vs. 0.754 g/L, p=0.003).

Discussion

Frequency of SERPINA1 gene variants in Switzerland.

The estimated mean gene frequencies for PiS and PiZ in Switzerland are 0.0384 and 0.0073, respectively[E4]. These data are based on the analysis of three Swiss cohorts[E5,E6] previously phenotyped for PI. The estimate is similar to ours for the S allele, but we obtained a slightly higher frequency for the Z allele. When we divided the population into the three language groups, we found evidence of a significantly higher frequency for the S allele in the French subgroup (p<0.001 for both comparisons, Table E1). This is in agreement with the hypothesis that the S mutation, which arose in the Portuguese population[E7], moved eastbound to the rest of Europe as a consequence of the late-glacial resettlement of Europe[E8], resulting in decreasing frequencies from southwest to the north and east.

AAT cut-off value for diagnostic purposes

One of the aims of this paper was to provide a clear cut-off, below which suspicion of AATD is reasonable, and to resolve the controversy around this issue. The choice of the AAT cut-off, below which samples should be selected for Pi pheno/genotyping, has important financial and clinical implications. The cut-offs determined by individual laboratories currently comprise a wide range between 1.00 and 1.30 g/L[E9-E11], and they strongly depend on specific requirements. For example, the clinical importance of PI*MS detection is considered far less important than PI*MZ detection, due to the different risks for emphysema for the two genotypes[E1,E2]. Therefore we reported two different cut-offs, one focused on avoiding the omission of deficient alleles S or Z (1.10 g/L) and the second identified genotypes at a likely risk for emphysema (1.00 g/L). We considered as well the previously referred cut-off of 1.13 g/L[E12], which is still useful since no Z alleles (and PI*SS) have been found in individuals with AAT blood levels above this level (100% sensitivity and 78.6%

specificity for detecting AATD genotypes), while 31% of all assigned PI*MS subjects show AAT blood levels higher than 1.13 g/L (78.6% sensitivity and 82.6% specificity for detecting any deficiency allele S or Z).

Effect of SERPINA1 M variants on AAT serum concentration.

The normal variants of AAT, usually called M, are characterized by point mutations that do not change either the phenotype or the serum concentration of the AAT protein. The most common are M1(Ala)/M1(Val) (213Ala, 213Val)[E13], M2 (213Val, 376Asp, and 101His)[E14], M3 (213Val and 376Asp)[E15], M4 (213Val, 101His)[E16]. It is well known that individuals bearing these mutations have normal AAT serum levels and that the protein functions normally as an inhibitor of neutrophil elastase. Evaluation of the crystallographic structure of AAT[E17] revealed that the substitution at position 101 occurs in helix D and the 376 substitution occurs in sheet 4B of the molecule, which are areas where changes are likely to cause minor conformational changes in the molecule.

Nevertheless, epidemiological studies of these normal variants are limited to their geographic distribution or prevalence data, and comparison among different genotypes, in terms of serum concentration of AAT, has never been performed. A side aim of this paper was to confirm the absence of quantitative differences among PI*MM subtypes; therefore, an analysis of all possible combinations of M1, M2, M3, and M4 alleles in the cluster PI*MM, was performed. The effect of the M1(Ala) allele, in combination with the Z allele, in reducing the mean concentration of AAT in serum (Figure E3), should be considered with caution.

Supplement References

E1. Hersh CP, Dahl M, Ly NP, *et al*. Chronic obstructive pulmonary disease in alpha 1-antitrypsin PI MZ heterozygotes: a meta-analysis. *Thorax* 2004; 59:843-9.

E2. Dahl M, Hersh CP, Ly NP, *et al*. The protease inhibitor PI*S allele and COPD: a meta-analysis. *Eur Respir J* 2005;26:67-76.

E3. Senn O, Russi EW, Schindler C, *et al*. Circulating alpha1-antitrypsin in the general population : determinants and association with lung function. *Respir Res* 2008;9:35

E4. de Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. *Chest* 2002;122:1818-29.

E5. Scheffrahn W, Ziggiotti E. Electrophoretic alpha 1-antitrypsin variation in the Swiss population. *Anthropol Anz* 1982;40:137-43.

E6. Bär W, Kratzer A. Polymorphism of alpha-1-antitrypsin (Pi) in the Swiss population determined by isoelectric focusing with an immobilized pH gradient. *Hum Hered* 1988;38:106-10.

E7. Seixas S, Garcia O, Trovoada MJ, *et al*. Patterns of haplotype diversity within the serpin gene cluster at 14q32.1: insights into the natural history of the alpha1-antitrypsin polymorphism. *Hum Genet* 2001;108:20-30.

E8. Pereira L, Richards M, Goios A, *et al*. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 2005;15:19-24.

E9. Barlow I, Sewell WA. Alpha1-antitrypsin deficiency and Pi typing. *J Allergy Clin Immunol* 2008;122:658.

E10. Corda L, Bertella E, Pini L, *et al*. Diagnostic flow chart for targeted detection of alpha1-antitrypsin deficiency. *Respir Med* 2006;100:463-70.

E11. De La Roza C, Rodrìguez-Frìas F, Lara B, *et al*. Results of a case-detection program for alpha-1 antitrypsin deficiency in COPD patients. *Eur Resp J* 2005; 26:616-22.

E12. Gorrini M, Ferrarotti I, Lupi A, *et al*. Validation of a rapid, simple method to measure alpha1-antitrypsin in human dried blood spots. *Clin Chem*. 2006;52:899-901.

E13. NukiwaT, Satoh K, Brantly ML, *et al*. Identification of a second mutation in the protein-coding sequence of the Z type alpha 1-antitrypsin gene. *J Biol Chem* 1986; 261:15989-94.

E14. Nukiwa T, Brantly ML, Ogushi F, *et al*. Characterization of the gene and protein of the common alpha 1-antitrypsin normal M2 allele. *Am J Hum Genet* 1988;43:322-30.

E15. Graham A, Hayes K, Weidinger S, *et al*. Characterisation of the alpha-1-antitrypsin M3 gene, a normal variant. *Hum Genet* 1990;85:381-2.

E16. Okayama H, Holmes MD, Brantly ML, *et al*. Characterization of the coding sequence of the normal M4 alpha 1-antitrypsin gene. *Biochem Biophys Res Commun* 1989;162:1560-70.

E17. Loebermann H, Tokuoka R, Deisenhofer J, *et al*. Human alpha1-proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J Mol Biol*. 1984;177:531-57.

Legends to Supplemental Figures

Figure E1. Intervals (5th-95th percentiles) for unadjusted serum AAT levels in the main *SERPINA1* genotypic classes; 1st-99th percentiles are represented with lines, where possible. In this figure, the g/L values of our analysis were mathematically converted to µM, based on a molecular weight of 52kDa.

Figure E2. ROC curve for predicting MM/MS vs. SS/MZ genotype classes from unadjusted AAT blood level (rare variants included).

Figure E3. Intervals (5th-95th percentiles) for AAT serum concentration in individuals stratified according to the haplotype reconstruction resulting in 24 genotypic classes.

Table E1. Frequencies of Z and S alleles in the cohort and the three language groups.

	Z allele (%)	S allele (%)
General population (n=6,057)	1.30	4.01
German subgroup (n=3,288)	1.22	3.65
French subgroup (n=1,938)	1.44	5.26
Italian subgroup (n=831)	1.26	2.53
p-value (chi-square-test)	0.61	<0.001

Table E2. Influence of gender and current smoking status on unadjusted AAT reference values.

	n	median	$5^{th}/95^{th}$ perc.
MM/MS all	5,848	1.26	1.01/1.63
MM/MS, male-smoker	791	1.29	1.01/1.59
MM/MS, female-smoker	663	1.36	1.07/1.74
MM/MS, male-nonsmoker	2,127	1.19	0.98/1.48
MM/MS, female-nonsmoker	2,267	1.28	1.03/1.69
SS/MZ, all	155	0.79	0.66/1.01
SS/MZ, male-smoker	16	0.79	0.69/1.02
SS/MZ, female-smoker	14	0.88	0.76/1.07
SS/MZ, male-nonsmoker	55	0.75	0.62/0.91
SS/MZ, female-nonsmoker	70	0.81	0.66/1.04

Table E3. Frequencies of the genotypic classes.

PI Group	PI Genotype	Frequencies
PI*MM		
	M1(Val) M1(Val)	0.2512
	M1(Ala) M1(Val)	0.1850
	M1(Val) M2 or M3 M4	0.1620
	M1(Val) M3	0.0908
	M1(Ala) M2	0.0559
	M1(Ala) M1(Ala)	0.0351
	M1(Ala) M3	0.0328

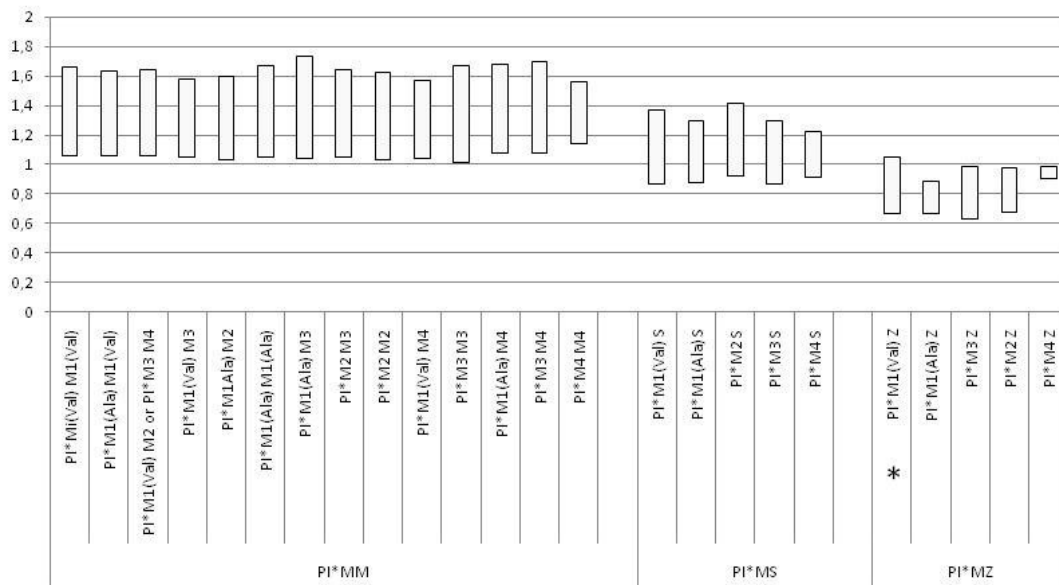|		
	M2 M3	0.0266
	M2 M2	0.0253
	M1(Val) M4	0.0113
	M3 M3	0.0105
	M1(Ala) M4	0.0060
	M3 M4	0.0042
	M4 M4	0.0007
PI*MS		
	M1(Val) S	0.0379
	M1(Ala) S	0.0173
	M2 S	0.0110
	M3 S	0.0078
	M4 S	0.0013
PI*MZ		
	M1(Val) Z	0.0133
	M1(Ala) Z	0.0047
	M3 Z	0.0028
	M2 Z	0.0027
	M4 Z	0.0003

Supplementary Figure 1

Supplementary Figure 2

Area under ROC curve = 0.9907
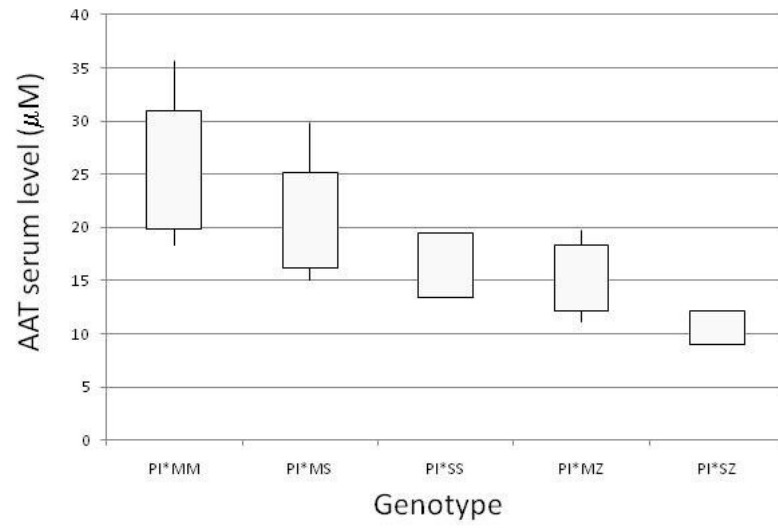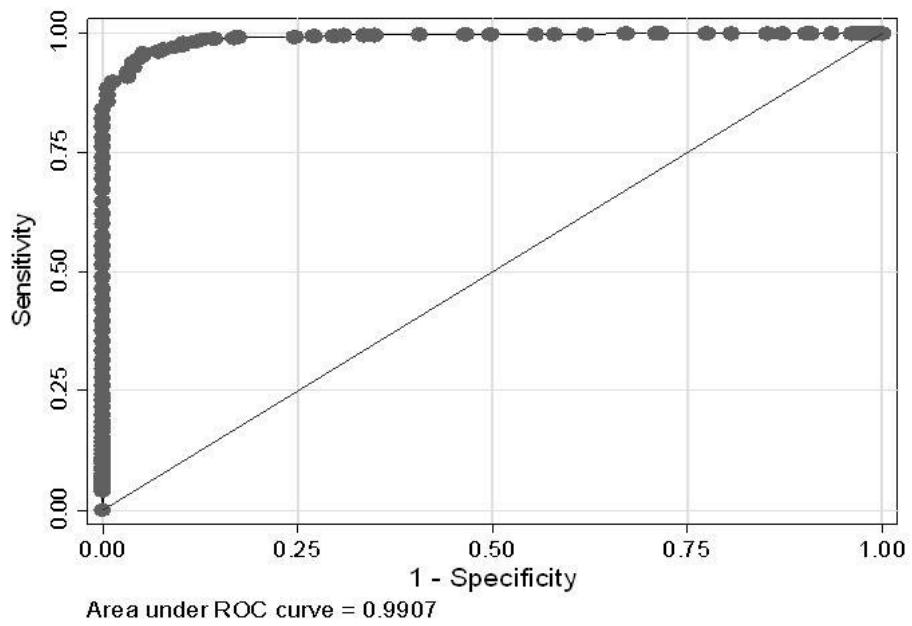
Supplementary Figure 3

* p=0.003; PI*M1(Val) Z vs PI*M1(Ala) Z (0.819 and 0.754 g/L, respectively)

5.2 Paper 3: Causal and Synthetic Associations of Variants in the *SERPINA* Gene Cluster with Alpha1-Antitrypsin Serum Levels.

This paper was published:

Thun GA, Imboden M, Ferrarotti I, Kumar A, Obeidat M, Zorzetto M, Haun M, Curjuric I, Couto Alves A, Jackson VE, Albrecht E, Ried JS, Teumer A, Lopez LM, Huffman JE, Enroth S, Bossé Y, Hao K, Timens W, Gyllensten U, Polasek O, Wilson JF, Rudan I, Hayward C, Sandford AJ, Deary IJ, Koch B, Reischl E, Schulz H, Hui J, James AL, Rochat T, Russi EW, Jarvelin M-R, Strachan DP, Hall IP, Tobin MD, Dahl M, Fallgaard Nielsen S, Nordestgaard BG, Kronenberg F, Luisetti M, Probst-Hensch NM. PLoS Genetics 2013; 9(8):e1003585.

The research article was accompanied by a perspective. Turner A. PLoS Genetics 2013; 9(8):e1003768.

Causal and Synthetic Associations of Variants in the *SERPINA* Gene Cluster with Alpha1-antitrypsin Serum Levels

Gian Andri Thun[1,2], Medea Imboden[1,2*], Ilaria Ferrarotti[3*], Ashish Kumar[1,2,4], Ma'en Obeidat[5], Michele Zorzetto[3], Margot Haun[6], Ivan Curjuric[1,2], Alexessander Couto Alves[7], Victoria E. Jackson[8], Eva Albrecht[9], Janina S. Ried[9], Alexander Teumer[10], Lorna M. Lopez[11], Jennifer E. Huffman[12], Stefan Enroth[13], Yohan Bossé[14], Ke Hao[15], Wim Timens[16], Ulf Gyllensten[13], Ozren Polasek[17], James F. Wilson[18], Igor Rudan[18], Caroline Hayward[12], Andrew J. Sandford[5], Ian J. Deary[11], Beate Koch[19], Eva Reischl[20], Holger Schulz[21], Jennie Hui[22,23,24], Alan L. James[25,26], Thierry Rochat[27], Erich W. Russi[28], Marjo-Riitta Jarvelin[7,29,30,31,32], David P. Strachan[33], Ian P. Hall[34], Martin D. Tobin[8], Morten Dahl[35], Sune Fallgaard Nielsen[36], Børge G. Nordestgaard[36], Florian Kronenberg[6], Maurizio Luisetti[3], Nicole M. Probst-Hensch[1,2*]

1 Swiss Tropical and Public Health Institute, Basel, Switzerland, 2 University of Basel, Basel, Switzerland, 3 Center for Diagnosis of Inherited Alpha1-antitrypsin Deficiency, Institute for Respiratory Disease, IRCCS San Matteo Hospital Foundation, University of Pavia, Pavia, Italy, 4 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, 5 James Hogg Research Centre, Institute of Heart and Lung Health, University of British Columbia, Vancouver, Canada, 6 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria, 7 Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom, 8 Departments of Health Sciences and Genetics, University of Leicester, Leicester, United Kingdom, 9 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, 10 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany, 11 Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom, 12 MRC Human Genetics, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, 13 Department of Immunology, Genetics, and Pathology, Rudbeck Laboratory, SciLifeLab, Uppsala University, Uppsala, Sweden, 14 Institut Universitaire de Cardiologie et de Pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec City, Canada, 15 Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, 16 Department of Pathology and Medical Biology, University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands, 17 Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia, 18 Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, United Kingdom, 19 Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany, 20 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, 21 Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, 22 School of Population Health, University of Western Australia, Perth, Australia, 23 Pathology and Laboratory Medicine, University of Western Australia, Perth, Australia, 24 Busselton Population Medical Research Foundation, Perth, Australia, 25 West Australian Sleep Disorders Research Institute, Perth, Australia, 26 School of Medicine and Pharmacology, University of Western Australia, Perth, Australia, 27 Division of Pulmonary Medicine, University Hospital of Geneva, Geneva, Switzerland, 28 Pulmonary Division, University Hospital of Zurich, Zurich, Switzerland, 29 Institute of Health Sciences, University of Oulu, Oulu, Finland, 30 Biocenter Oulu, University of Oulu, Oulu, Finland, 31 Unit of Primary Care, Oulu University Hospital, Oulu, Finland, 32 Department of Children and Young People and Families, National Institute for Health and Welfare, Oulu, Finland, 33 Division of Population Health Sciences and Education, St George's, University of London, London, United Kingdom, 34 Division of Therapeutics and Molecular Medicine, Queen's Medical Centre, Nottingham, United Kingdom, 35 Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark, 36 Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark

Abstract

Several infrequent genetic polymorphisms in the *SERPINA1* gene are known to substantially reduce concentration of alpha1-antitrypsin (AAT) in the blood. Since low AAT serum levels fail to protect pulmonary tissue from enzymatic degradation, these polymorphisms also increase the risk for early onset chronic obstructive pulmonary disease (COPD). The role of more common *SERPINA1* single nucleotide polymorphisms (SNPs) in respiratory health remains poorly understood. We present here an agnostic investigation of genetic determinants of circulating AAT levels in a general population sample by performing a genome-wide association study (GWAS) in 1392 individuals of the SAPALDIA cohort. Five common SNPs, defined by showing minor allele frequencies (MAFs) >5%, reached genome-wide significance, all located in the *SERPINA* gene cluster at 14q32.13. The top-ranking genotyped SNP rs4905179 was associated with an estimated effect of $\beta = -0.068$ g/L per minor allele ($P = 1.20 \times 10^{-12}$). But denser *SERPINA1* locus genotyping in 5569 participants with subsequent stepwise conditional analysis, as well as exon-sequencing in a subsample (N = 410), suggested that AAT serum level is causally determined at this locus by rare (MAF<1%) and low-frequent (MAF 1–5%) variants only, in particular by the well-documented protein inhibitor S and Z (PI S, PI Z) variants. Replication of the association of rs4905179 with AAT serum levels in the Copenhagen City Heart Study (N = 8273) was successful (P<0.0001), as was the replication of its synthetic nature (the effect disappeared after adjusting for PI S and Z, P = 0.57). Extending the analysis to lung function revealed a more complex situation. Only in individuals with severely compromised pulmonary health (N = 397), associations of common SNPs at this locus with lung function were driven by rarer PI S or Z variants. Overall, our meta-analysis of lung function in ever-smokers does not support a functional role of common SNPs in the *SERPINA* gene cluster in the general population.

Introduction

Alpha1-antitrypsin (AAT) is a serum marker for inflammation produced in the liver. Its main function is to inhibit neutrophil elastase and consequently protect pulmonary tissue. The *SERPINA1* gene encoding the AAT protein is known to be polymorphic in the general population. The best studied single nucleotide polymorphisms (SNPs) causing a reduction in AAT serum levels are the protease inhibitor S (PI S, rs17580) and the protease inhibitor Z (PI Z, rs28929474) variants [1]. The loss of function mechanism is especially well investigated for the PI Z variant. The resulting amino acid change in AAT leads to the protein's intracellular polymerization in hepatocytes and therefore to a reduced level of secreted serum AAT [2]. Homozygosity for

PI Z (PI ZZ genotype) with a frequency of about 0.01% in Caucasian populations [3] causes blood AAT levels below 30% of normal. This genotype is clearly associated with elevated chronic obstructive pulmonary disease (COPD) risk accounting for 1–2% of all cases [4,5]. There is also strong evidence that accelerated lung function decline and increased obstructive disease risk can be caused by compound heterozygosity of PI Z and PI S (PI SZ genotype). The case is less clear for PI MZ, PI MS or PI SS genotypes (PI M standing for the normal allele), which cause a less pronounced reduction in AAT concentration, as previous studies produced inconsistent evidence [6–9].

Further of note, large-scale genome-wide association studies (GWAS) on COPD or on cross-sectional or longitudinal lung function have not identified the *SERPINA1* gene to be a major

Author Summary

Low levels of alpha1-antitrypsin (AAT) in the blood are a well-established risk factor for accelerated loss in lung function and chronic obstructive pulmonary disease. While a few infrequent genetic polymorphisms are known to influence the serum levels of this enzyme, the role of common genetic variants has not been examined so far. The present genome-wide scan for associated variants in approximately 1400 Swiss inhabitants revealed a chromosomal locus containing the functionally established variants of AAT deficiency and variants previously associated with lung function and emphysema. We used dense genotyping of this genetic region in more than 5500 individuals and subsequent conditional analyses to unravel which of these associated variants contribute independently to the phenotype's variability. All associations of common variants could be attributed to the rarer functionally established variants, a result which was then replicated in an independent population-based Danish cohort. Hence, this locus represents a textbook example of how a large part of a trait's heritability can be hidden in infrequent genetic polymorphisms. The attempt to transfer these results to lung function furthermore suggests that effects of common variants in this genetic region in ever-smokers may also be explained by rarer variants, but only in individuals with hampered pulmonary health.

genetic determinant [10–12]. But a recent GWAS on emphysema [13] and a comprehensive evaluation of candidate regions for lung function [14] reported rs4905179 and rs3748312, two common SNPs (minor allele frequencies (MAFs) >5%) located in the *SERPINA* gene cluster on 14q32.13, among their most strongly associated results. This locus encompasses *SERPINA1* and ten other genes (*SERPINA2* to *SERPINA6* and *SERPINA9* to *SERPINA13*) encoding extracellular 'clade A' serpins with very heterogeneous functions [15]. It is currently not known whether such association signals observed for this locus reflect a causal role of common variants or whether they are merely synthetic, reflecting effects of rarer causal variants [16]. Towards that aim, but also to detect further chromosomal loci of potential relevance to circulating levels of AAT, we first performed a GWAS on AAT serum level using a subset of the population-based Swiss Cohort Study of Air Pollution and Lung Disease in Adults (SAPALDIA) as discovery sample, and a second subset of SAPALDIA as well as an independent cohort, the Copenhagen City Heart Study (henceforth referred to as Copenhagen), as replication sample. We also conducted fine mapping analyses of the *SERPINA1* gene in the SAPALDIA cohort. Finally, we meta-analyzed the lung function effect of common and low-frequent *SERPINA1* SNPs previously observed to be associated with pulmonary health in ever-smokers, based on data provided by several population- and patient-based studies.

Results

The discovery population and the design used to determine AAT-associated genetic variants are depicted in Figure S1 and further described in the Materials and Methods section. A comparison between the characteristics of the genome-wide analyzed sample (SAPALDIA discovery arm, N = 1392) and the remainder of the SAPALDIA cohort (SAPALDIA replication arm, N = 4245) did not reveal substantial differences in AAT serum levels or covariate distribution (Table S1), although asthmatics

were overrepresented (39.4%) in the SAPALDIA discovery arm and absent in the replication arm, which is due to previous study design [17]. The participants of the independent replication cohort Copenhagen (N = 8273) were on average five years older and had twice as many current smokers (Table S1). This was in line with substantially lower lung function levels (more than 800 mL lower forced expiratory volume in 1 second, FEV1, compared to both SAPALDIA subsets) and slightly elevated AAT blood levels (1.339 g/L vs. 1.257 and 1.255 g/L, respectively). The characteristics of the study populations contributing to the genetic association analyses with lung function are given in Table S2.

GWAS on AAT Serum Level

The association of more than 2.1 million genome-wide SNPs with AAT serum levels is shown in Figure 1. The ten most strongly associated SNPs were all located in the *SERPINA* gene cluster, half of them reached genome-wide significance ($P < 5*10^{-8}$, Table 1). The top 100 ranking SNPs are provided in Table S3. A regional association plot for the *SERPINA* gene cluster is shown in Figure 2. Both the top-ranking imputed SNP, rs2736887, and the top-ranking genotyped SNP, rs4905179, were located in close proximity to the *SERPINA6* gene and approximately 33 kb and 50 kb downstream of *SERPINA1* (effect estimates $\beta = -0.071$ and -0.068 g/L per minor allele; $P = 2.48*10^{-13}$ and $1.20*10^{-12}$, respectively). Linkage disequilibrium (LD) between these two variants based on HapMap2 CEU (Utah residents with Northern and Western European ancestry) derived haplotype data [18] was strong ($r^2 = 0.88$, $D' = 1$), but Figure 2 suggests that the LD, expressed in r^2 values, between the top-ranking SNP and the other SNPs in the region is generally modest. The genomic inflation factor lambda was low ($\lambda = 1.02$), suggesting minimal population stratification. The quantile-quantile plot (Q-Q plot) showed good adherence to null expectation and substantial positive deviation between observed and expected p-values for the top-ranking SNPs (Figure S2). In a sensitivity analysis adjusting for additional covariates, including high sensitivity C-reactive protein (hs-CRP), body mass index (BMI), passive smoking and alcohol intake, the genome-wide association results did not show an increase in the strength of the top-ranking loci, nor did they point to additional loci (data not shown). Even though this GWAS was enriched with asthma patients, GWAS stratification according to asthma status did not show heterogeneity for the top-ranking signals between participants with and without asthma (data not shown).

Association of 1000 Genomes Imputed Data for the *SERPINA* Gene Cluster with AAT Serum Level

In order to further refine association signals in this region, we imputed additional SNPs on chromosome 14 using haplotype data from the 1000 Genomes Project (1000G) [19]. The 1000G imputation yielded a three times higher number of imputed variants with reasonable quality scores (imputation-$r^2 > 0.5$) compared to HapMap-derived imputed variants. A region defined by 1 Mb up- and downstream of the *SERPINA1* gene revealed 24 additional variants that were associated below a local significance level of $P < 3*10^{-5}$, adjusting for approximately 1800 SNPs covering a region of 2 Mb (Table S4). Among them, four low-frequent variants and one rare variant showed p-values reaching genome-wide significance level, and interestingly, none of them was in high LD ($r^2 > 0.8$) with any other regional variant tested. The most strongly associated signal came from the PI Z variant, which is well known to be associated with reduced AAT serum levels ($\beta = -0.620$ g/L per minor allele, $P = 4.61*10^{-43}$, MAF = 0.84%). The other well established
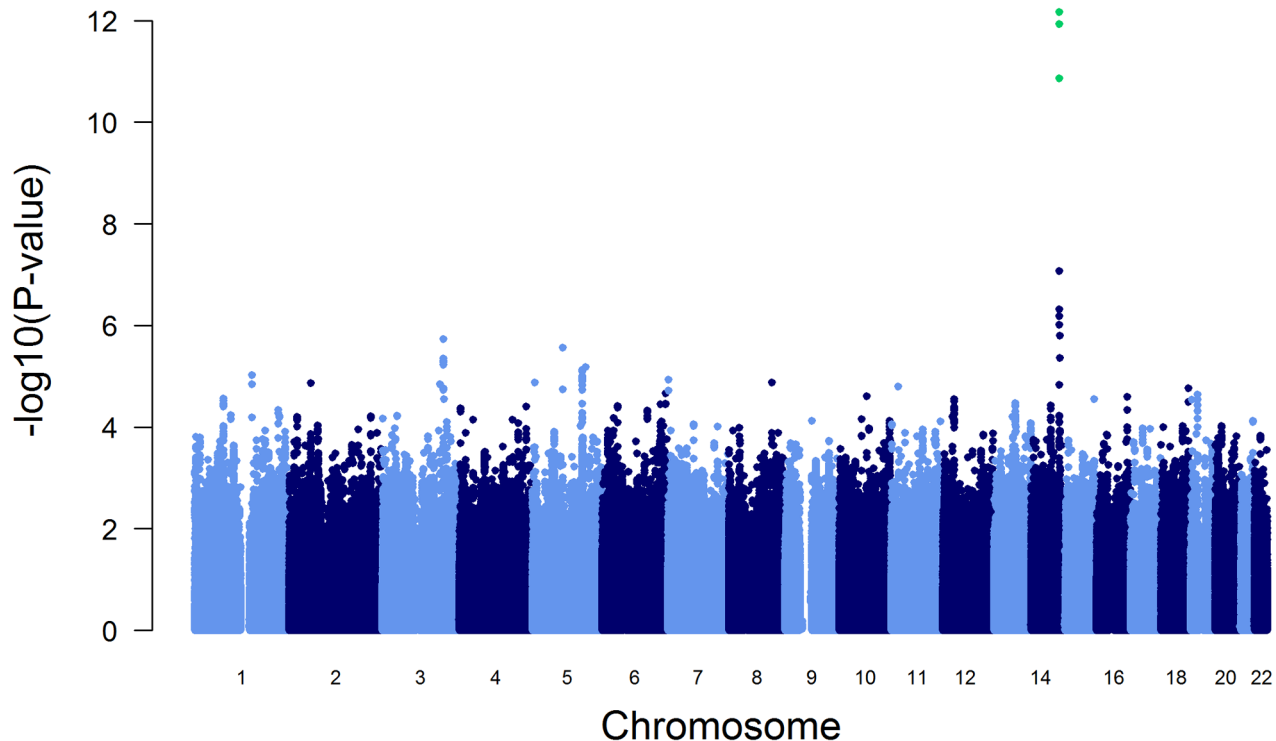
Figure 1. Manhattan plot of genome-wide -log(10) p-values for association with AAT serum level. SNPs reaching genome-wide significance are shown in green. They all belong to the *SERPINA* gene cluster.
doi:10.1371/journal.pgen.1003585.g001

causal polymorphism, the PI S variant, was less prominently ranked ($\beta = -0.110$ g/L per minor allele, $P = 1.95 \times 10^{-6}$, MAF = 5.70%) and exhibited an insufficient imputation quality (imputation-$r^2 = 0.45$).

GWAS on AAT Serum Level, Conditional on PI S and PI Z Variants

Accuracy of the imputed PI S and PI Z results was confirmed by direct genotyping of the samples [20]. The discovery arm revealed

33 PI Z carriers, 111 PI S carriers and two compound heterozygous carriers of PI S and PI Z (MAF = 1.26% for PI Z and 4.06% for PI S, respectively). No homozygous PI S or PI Z genotypes were detected.

To test the influence of these variants on the initially reported GWAS results (Figure 1), we performed a conditional GWAS by additionally adjusting the regression models for the presence of PI S and PI Z alleles. We observed a drastic change in the association of the *SERPINA* gene cluster SNPs with AAT serum

Table 1. The ten most strongly associated SNPs in the unconditional GWAS on AAT serum level in SAPALDIA (N = 1392).

SNP	Chromosome	Position	Gene	Location	Determination	MAF	Imp-r²	Allele Effect	P
rs2736887	14	93882733		intergenic	imputed	0.185	0.950	0.071	2.48E-13
rs926144	14	93883155		intergenic	imputed	0.186	0.950	0.071	2.72E-13
rs7151526	14	93933389	*SERPINA1*	5'UTR	imputed	0.065	0.769	0.116	6.78E-13
rs4905179	14	93865245	*SERPINA6*	5'UTR	genotyped	0.180	1.000	0.068	1.20E-12
rs11621961	14	93839229	*SERPINA6*	3'UTR	genotyped	0.355	0.945	0.052	1.37E-11
rs17751837	14	93937997	*SERPINA1*	5'UTR	genotyped	0.097	0.995	0.063	8.56E-08
rs1028580	14	93919635	*SERPINA1*	intron	imputed	0.154	0.979	0.051	4.87E-07
rs8010121	14	93920367	*SERPINA1*	intron	genotyped	0.155	0.999	0.049	6.64E-07
rs3748312	14	93924017	*SERPINA1*	intron	imputed	0.148	0.846	0.053	9.84E-07
rs17752593	14	94007781	*SERPINA9*	intron	genotyped	0.129	0.997	0.053	1.59E-06

Abbreviations: AAT, alpha1-antitrypsin; GWAS, genome-wide association study; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Imp-r^2 is an indicator for imputation quality. SNPs with MAF<0.05 or imp-r^2<0.5 were excluded.
Chromosomal position is based on reference panel, NCBI build 36.3. Allele effects are shown in absolute numbers.
doi:10.1371/journal.pgen.1003585.t001
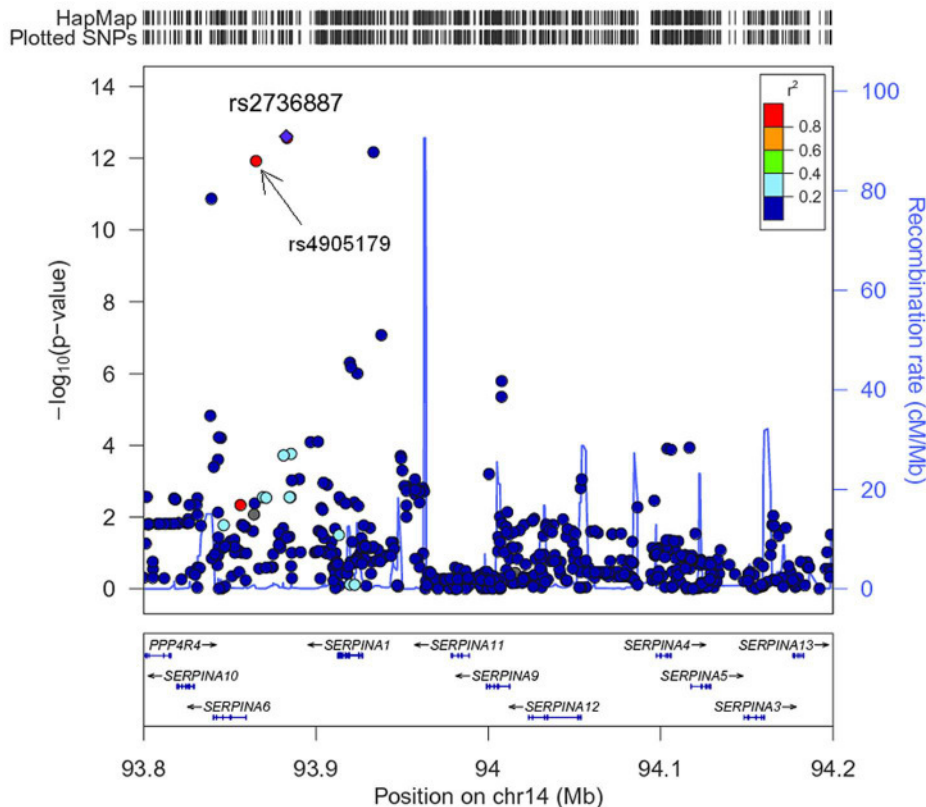
Figure 2. Regional plot for the *SERPINA* gene cluster (93.8–94.2 Mb on chromosome 14q32.13, reference panel: NCBI build 36.3). Presented are -log(10) p-values and LD (r^2) with top-ranking SNP rs2736887 (purple diamond) for all SNPs in this region. The blue line shows recombination rate.
doi:10.1371/journal.pgen.1003585.g002

level (Figure 3 and Table 2). The strong signal on chromosome 14 observed in the original GWAS disappeared completely and the top-ranking imputed and genotyped SNPs (rs2736887 and rs4905179) were no longer significant (P = 0.44 and 0.31, respectively). In fact, no SNP was found near the *SERPINA* gene cluster among the 100 most strongly associated common variants (Table S5). In addition, the 1000G imputed data, comprising sequences 1 Mb up- and downstream of the *SERPINA1* gene, did not show evidence of other independent AAT-associated SNPs. An alternative approach that excluded all PI S and PI Z carriers from the GWAS sample (N = 146), instead of adjusting for them, confirmed the results. Both analyses revealed an intergenic region on chromosome 3 with borderline genome-wide significance (top-ranking SNP rs2566347, β = −0.043 g/L per minor allele, P = 7.88*10^{-8}, in the adjusted GWAS). The top SNPs in this region were located in proximity to *MFSD1* and *RARRES1*, which are two genes with sparsely annotated function. The 1000G imputation of this region did not reveal further variants. In addition, we were unable to replicate this association signal in the SAPALDIA replication arm (N = 4245, β = −0.004 g/L per minor allele, P = 0.46).

Replication in the Copenhagen City Heart Study

The effect of rs4905179, the top genotyped SNP in our GWAS, on AAT serum levels was tested for replication in Copenhagen (Table 3). The minor allele was associated with β = −0.097 g/L (P<0.0001, N = 8332). As observed in the GWAS, adjustment for PI S and Z polymorphisms resulted in a complete loss of this signal (β = 0.003 g/L, P = 0.57, N = 8273).

Impact of Common and Low-Frequent *SERPINA1* Genetic Variants on AAT Serum Level

In a first fine mapping step, 16 *SERPINA1* SNPs (see Materials and Methods section for a description of the SNP selection) were successfully genotyped in 5569 SAPALDIA subjects (discovery and replication arm combined). The genotype results in the discovery arm allowed us to compare allele frequencies with imputed results derived from the 1000G data. Table S6 shows that the agreement was very high. Stepwise conditional regression analyses were then applied to evaluate the independent effects of each of these SNPs on AAT serum levels (Table 4). The PI Z variant was most strongly associated with circulating levels of AAT. The PI S variant remained strongly associated after conditioning on PI Z. Two variants located in the 5′ non-coding gene region (rs2896268 and rs1956707) were marginally associated with the phenotype in two further steps after conditioning on PI S and PI Z. The total variance of AAT explained by statistical models increased from 8.8% (model with only non-genetic factors) to 32.6% (adding PI S and PI Z alleles), and to 32.8% adding rs2896268 and rs1956707. Based on genotype data from the SAPALDIA cohort, the *SERPINA1* gene contains three haplotype blocks using D′-based block definition (Figure 4). The AAT deficiency variants PI S and PI Z are located in block 1, while rs2896268 and rs1956707 are located in block 3, roughly 8 kb upstream of exon 1.

Impact of Rare *SERPINA1* Genetic Variants on AAT Serum Level

In a second fine mapping step, exon sequencing was performed in 410 subjects with low AAT levels that were independent of the
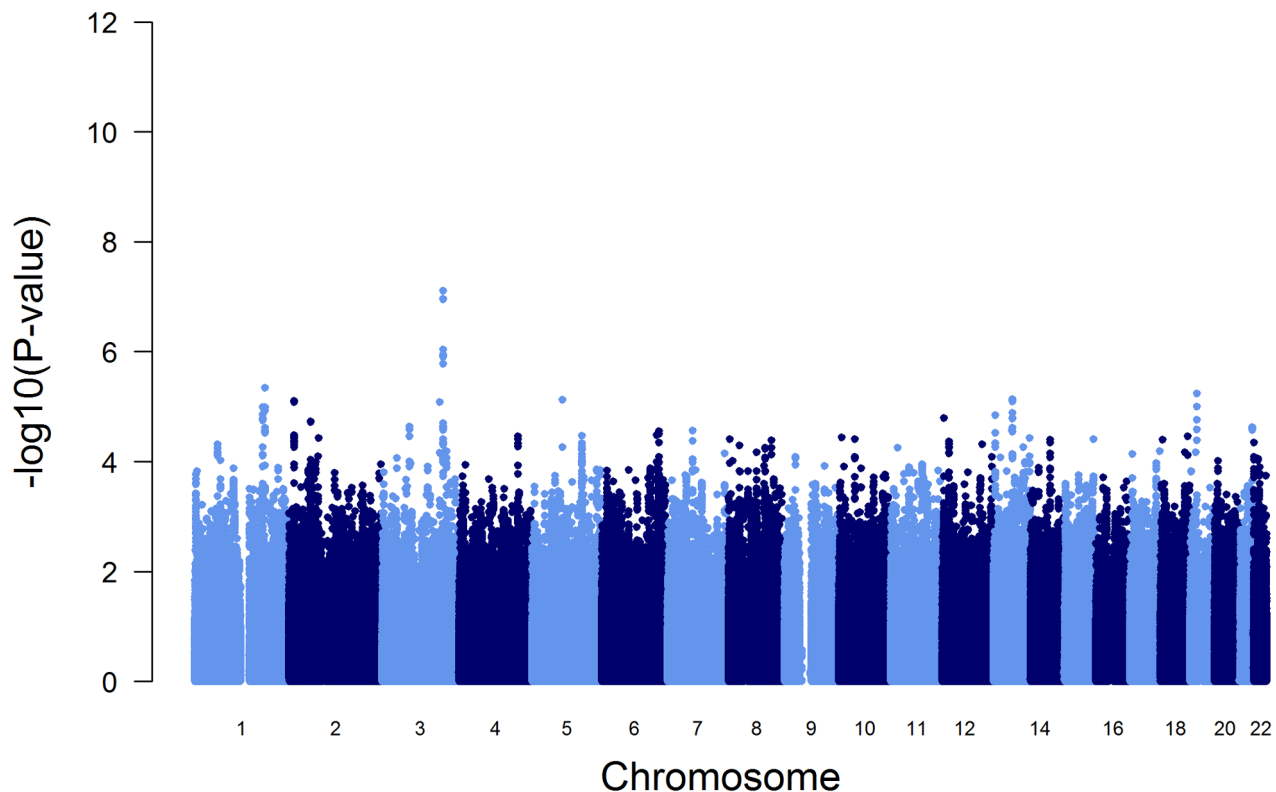
Figure 3. Manhattan plot of genome-wide -log(10) p-values for association with AAT serum level, conditional on PI S and PI Z alleles.
doi:10.1371/journal.pgen.1003585.g003

presence of PI S or PI Z alleles [21]. 16 additional *SERPINA1* variants (two deletions and 14 SNPs) were detected, of which all but one had already been described [21–29] (Table S7). Three of the SNPs were synonymous, and five had no accession numbers in public databases (as of April 1ˢᵗ, 2013). Most of the non-synonymous SNPs have already been described as potentially lowering AAT serum level, and computational tools only classified one of them as no damaging to the protein's tertiary structure. In order to estimate the phenotypic influence of these rare variants, we compared mean AAT blood levels, adjusted for sex, age, study center, current smoking, as well as for the presence of PI S and Z alleles, between samples without rare variants (N = 346) and those with a single rare variant (N = 63) or more than one (N = 1). The subjects with rare variants had a lower adjusted mean AAT level (0.904 g/L, 95% CI 0.884 to 0.924 g/L) compared to those without rare variants (0.992 g/L, 95% CI 0.984 to 1.000 g/L, P<0.001). Although this difference is small, the range covers the recently proposed upper limit of intermediate AAT deficiency (0.92 g/L), a value with some clinical relevance [20]. AAT levels of carriers of synonymous mutations or non-synonymous mutations without predicted damaging consequences to protein structure (N = 20) were not different from those carrying no rare variants (0.985 vs. 0.990 g/L, P = 0.77). Assuming that unsequenced samples were negative for mutations with predicted deleterious functional effects, the total variance of explained AAT further increased from 32.8% to 35.4% (based on a statistical model adding all rare mutations with predicted damaging consequences to the protein structure).

Common and Low-Frequent *SERPINA1* SNPs Previously Associated with Lung Function

Results from a previous GWAS on emphysema [13] and a large-scale evaluation of candidate loci on lung function [14] pointed to a role of common variants in the *SERPINA* gene cluster. The SNPs rs4905179 (associated with emphysema in smokers [13]) and rs3748312 (associated with cross-sectional lung function among ever smokers [14]) were strongly associated with AAT in our study (Tables 1 and 4), but both signals disappeared upon adjustment for the low-frequent variants PI S and Z. In order to clarify whether the association of the two common SNPs with pulmonary health could also be explained by effects of the rarer SNPs, we conducted a meta-analysis for cross-sectional lung function in ever-smokers across 17 studies with a total sample size of N = 24,446 (Table S2). We included nine studies which had contributed to the original finding on lung function [14] and had available genotypes or 1000G imputed genotype data on PI S and Z. The meta-analysis in cohorts of general population study design showed that rs4905179 was not associated with lung function in ever-smokers (P = 0.90 in the fixed-effect meta-analysis, N = 20,153, Figure 5). Yet smaller studies recruited within population isolates showed a trend for the rare allele to be associated with low lung function (random-effect P = 0.02, N = 1623, Figure 5), and in contrast to the association with AAT serum levels, adjusting for PI S and Z alleles did not modify the association of rs4905179 with lung function (Figure 6). For the second common SNP, rs3748312, we could nominally replicate the statistically significant allele effect on FEV1 in the general population of ever-smokers (P = 0.02, N = 15,450), and the

Table 2. The ten most strongly associated SNPs in the GWAS on AAT serum level, conditional on PI S and PI Z alleles in SAPALDIA (N = 1392).

SNP	Chromosome	Position	Gene	Location	Determination	MAF	Imp-r^2	Allele Effect	P
rs2566347	3	159974071		intergenic	imputed	0.192	0.998	0.043	7.88E-08
rs1560417	3	159972476		intergenic	imputed	0.200	0.998	0.042	1.11E-07
rs1560418	3	159972335		intergenic	genotyped	0.200	1.000	0.042	1.11E-07
rs1430414	3	159987697	*MFSD1*	5′UTR	imputed	0.137	0.984	0.045	9.26E-07
rs6761989	3	159983253		intergenic	imputed	0.137	0.993	0.044	1.14E-06
rs17643917	3	159968433		intergenic	imputed	0.137	1.000	0.044	1.23E-06
rs17643860	3	159967954		intergenic	imputed	0.137	1.000	0.044	1.24E-06
rs17700475	3	159967627		intergenic	genotyped	0.137	1.000	0.044	1.25E-06
rs3863076	3	159969394		intergenic	genotyped	0.145	1.000	0.042	1.69E-06
rs2206593	1	184909052	PTGS2	3′UTR	genotyped	0.065	0.956	0.060	4.60E-06

Abbreviations: AAT, alpha1-antitrypsin; GWAS, genome-wide association study; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Imp-r^2 is an indicator for imputation quality. SNPs with MAF<0.05 or imp-r^2<0.5 were excluded.
Chromosomal position is based on reference panel, NCBI build 36.3. Allele effects are shown in absolute numbers.
doi:10.1371/journal.pgen.1003585.t002

stronger effect that was published [14] seems to be driven by population isolates (Figure 7). Again, as for rs4905179, the associations were not dependent on S and Z alleles (Figure 8). Meta-analyses of the associations of PI S and Z alleles with lung function revealed no consistent associations between these functional AAT level determining variants and reduced FEV1 (Figures S3 and S4). Remarkably, the significant associations of rs4905179 and rs3748312 with lung function assessed in two additional studies with patients featuring compromised pulmonary health and undergoing lung resection, showed evidence for synthetic associations of the common SNPs with lung function that are consistent with our results for circulating AAT (Table 5). The minor alleles were associated with lower lung function and the association completely disappeared when conditioned on the presence of PI S and Z alleles.

Discussion

We present here the first GWAS on circulating AAT blood levels. Our results confirm that genetic variation in the *SERPINA1* gene is a strong determinant of serum AAT levels. Fine mapping of *SERPINA1* and subsequent stepwise regression analyses further revealed that the associations with common variants in the *SERPINA* locus could be attributed to rarer variants previously identified to be causally linked with AAT deficiency.

There is an ongoing debate about whether rare variants are responsible for the missing heritability observed in GWAS on many complex outcomes [30]. We show here an example in which

the polymorphisms PI S and PI Z seem to account for basically all observable effects of common variants in the *SERPINA* gene cluster on AAT serum level. The top-ranking genotyped SNP in our GWAS, rs4905179, was in low r^2-based LD with PI S (r^2 = 0.18) and PI Z (r^2 = 0.06), reflecting in part the unequal allele frequencies of these SNPs. However, PI S and PI Z showed very high LD in terms of D′ with the GWAS top signals (e.g. D′ = 0.95 and 0.96, respectively, with rs4905179) and generally with many common variants in this locus (Figure 4), suggesting little genetic recombination. This proof-of-principle approach, revealing that signals of common variants in fact merely reflect rarer variants, has recently also been shown for some of the loci regulating low-density lipoprotein (LDL) cholesterol [31,32]. Yet for other loci linked to LDL cholesterol, as well as for loci influencing other traits, both common and low-frequent variants contributed independently of the original GWAS signal to the phenotypic trait [31,33,34].

Using regional 1000G imputation within the top-ranking loci can allow the identification of additional association signals of stronger size to support the initial GWAS top result, as observed here for the *SERPINA* cluster, but not for the locus near *MFSD1*, an association which was not confirmed in the SAPALDIA replication arm. The resequencing strategy of the GWAS-identified locus in a sample with low AAT concentrations yielded in the identification of rare variants being strongly associated with reduced AAT blood levels. Such an accumulation of rare variants in the extreme range of the respective phenotype has also been reported by others [35,36]. As for the relative contribution of

Table 3. Minor allele effects of PI S, PI Z and rs4905179 on AAT serum levels in the Copenhagen City Heart Study.

SNP	N	MAF (genotyped)	Allele Effect (g/L)	95% Confidence Intervals	P
rs17580 (PI S)	8338	0.029	−0.188	−0.211 to −0.165	<0.0001
rs28929474 (PI Z)	8338	0.027	−0.492	−0.514 to −0.470	<0.0001
rs4905179	8332	0.186	−0.097	−0.107 to −0.087	<0.0001
rs4905179, adjusted for PI S and Z	8273	0.186	0.003	−0.007 to 0.013	0.57

Abbreviations: MAF, minor allele frequency; SNP, single nucleotide polymorphism.
doi:10.1371/journal.pgen.1003585.t003

Table 4. Common and low-frequent *SERPINA1* SNPs and their association with AAT serum level, univariate and conditional on significantly associated SNPs (N = 5569[a]), in SAPALDIA.

SNP	Location	Position	Selection[b]	MAF	Univariate[c]		Conditional[d]	
					Allele Effect	P	Allele Effect	P
rs2896268	5′UTR	93935461	C	0.495	0.006	0.10	**0.013**	**4.1E-05**
rs1956707	5′UTR	93933946	C	0.038	0.016	0.10	**0.029**	**5.0E-04**
rs8004738	exon 1	93926667	D	0.490	0.005	0.15	0.001	0.81
rs1570142	intron 1	93926015	A,B,C	0.488	0.005	0.19	0.001	0.88
rs3748312	intron 1	93924017	A,B	0.153	0.035	2.4E-12	0.002	0.70
rs3748316	intron 1	93923617	A,C	0.181	0.011	0.02	0.002	0.63
rs3748317	intron 1	93923432	A	0.158	0.020	5.5E-05	0.004	0.36
rs1980617	intron 1	93922287	A	0.389	0.030	5.0E-16	0.003	0.31
rs1980618	intron 1	93922176	A,C	0.383	0.030	1.3E-16	0.005	0.15
rs2753935	intron 1	93920690	A	0.435	0.012	9.7E-04	0.004	0.24
rs2144831	intron 1	93919723	A,C	0.240	0.021	9.8E-07	0.002	0.71
rs709932	exon 2	93918954	A,C	0.169	0.021	1.2E-05	0.003	0.50
rs6647	exon 3	93917168	A,C	0.200	0.026	7.2E-09	0.003	0.51
rs17580 (PI S)	exon 3	93917015	D	0.041	0.210	1.9E-127	**0.218**	**1.9E-163**
rs28929474 (PI Z)	exon 5	93914700	D	0.013	0.483	1.3E-240	**0.482**	**5.2E-269**
rs1303	exon 5	93914596	A,C	0.247	0.015	4.7E-04	0.001	0.89

Abbreviations: AAT, alpha1-antitrypsin; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Chromosomal position is based on reference panel, NCBI build 36.3.
[a]Includes subjects for whom all the 16 SNPs have been successfully genotyped.
[b]SNP selection was based on extreme trait sequence data (A), tagging SNPs according to HapMap (B), TAMAL software (C) and publication about functionality (D); see Materials and Methods for a more detailed description.
[c]Univariate analyses were adjusted for non-genetic factors only (sex, age, recruiting area and current smoking status). Allele effects are shown in absolute numbers and P<0.005 was considered statistically significant.
[d]In a forward selection approach of stepwise regression, the four SNPs in bold contributed statistically significantly to the variability in AAT serum levels and were included in the final statistical model. Allele effects and p-values refer to this final model.
doi:10.1371/journal.pgen.1003585.t004

genetic variants on the phenotype, we confirmed that effect sizes of PI S and Z on AAT serum levels were comparably strong, explaining alone a high proportion of the total variability (24.2%). We estimated that rare variants explained at least another 2% in our population-based sample, but since we did not sequence the entire SAPALDIA sample for rare variants, we cannot reliably quantify this contribution. In terms of blood markers, similar examples exist in which one genetic variant could explain well above 5% of the phenotype's variability (e.g. lipoprotein(a) [37], bilirubin [38] or adiponectin [39]), but for many other markers like serum lipid levels, only variants with small effects have been detected so far [40].

Association patterns between *SERPINA1* variants and circulating AAT did not translate to according associations with lung function level in a straightforward manner. Lung function is a complex phenotype associated with numerous genetic variants [11,12]. Studies on the associations of *SERPINA1* polymorphisms with lung function and COPD have produced mixed results. It is well accepted that severe AAT deficiency caused by PI null mutations or by the presence of two PI Z alleles puts a subgroup of carriers at higher risk of emphysema and COPD, especially when smoking [5]. Studies on COPD found suggestive evidence for an association with heterozygous status for the PI Z allele [6,8], but we observed no associations in our meta-analysis between PI Z and lung function level in ever-smokers. In the SAPALDIA general population sample, we had previously reported that an effect of the PI Z allele on lung function decline is restricted to

persistent smokers and primarily observed for forced expiratory flow 25–75% [9]. Other variation in or close to the *SERPINA1* gene has been proposed to play a role for pulmonary health. First, a haplotype pattern of five common SNPs was reported to be more frequent in COPD cases than in controls in a study with limited statistical power [41]. The only SNP which was also separately associated with COPD in that analysis was not associated with reduced serum levels in our study (rs8004738, Table 4). Second, the minor allele of rs4905179, which was the top signal in the current AAT GWAS, was positively associated with emphysema assessed by chest tomography in three independent cohorts consisting of smoking COPD patients without severe AAT deficiency (PI ZZ) [13]. Finally, the minor allele of the intronic SNP rs3748312 was positively associated with lung function in ever-smokers from different population-based studies of the SpiroMeta Consortium [14]. The association of these two SNPs with lung function in ever-smokers was heterogeneous across studies in our meta-analysis. Dependency on PI S and Z was limited to studies in patients with lung resection (Groningen, UBC), consistent with the notion that *SERPINA1* may only confer risk in selected population subgroups.

There are several possible explanations for the poor translation of genetic association patterns with serum AAT to lung function and for the heterogeneity of associations between *SERPINA1* variants and lung function. First, lung function is influenced by mechanisms in addition to protease-antiprotease disequilibrium. Second, the contribution of the *SERPINA1* gene variants as a
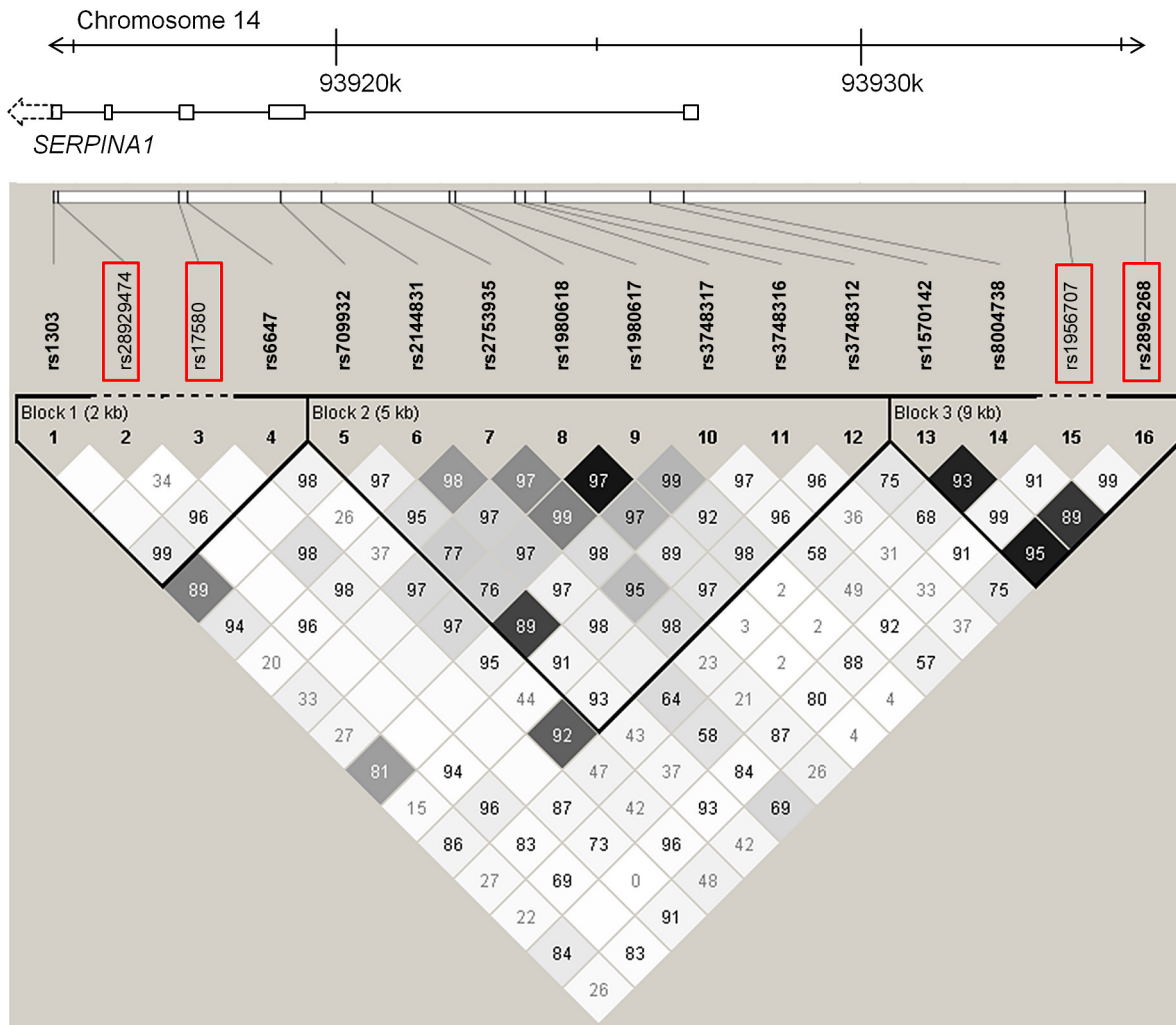
Figure 4. LD plot among common and low-frequent SNPs in the *SERPINA1* gene within the SAPALDIA study. Shading represent r[2] values, whereas numbers represent D′ values (no number equals D′ = 1). Red framed SNPs are independently associated with AAT serum levels after forward selection stepwise regression modeling. Rs17580 is the PI S variant and rs28929474 is the PI Z variant.
doi:10.1371/journal.pgen.1003585.g004

determinant of lung function likely depends on both the evolutionary pressure in isolated populations and the prevalence of effect modifiers in the respective study populations. These include smoking and smoking intensity, and likely other markers of inflammation. AAT itself plays a dual role in its relationship with lung function. While chronic AAT deficiency is etiologically associated with adverse pulmonary health, individuals with lung function impairment in fact exhibit higher AAT levels for a given genetic background due to AAT's role as an acute-phase inflammation marker [42,43]. Third, tissue-specific regulation of the *SERPINA1* locus may play an important role. Serum AAT levels are driven by *SERPINA1* expression, protein formation and secretion in hepatocytes, so that regulatory SNPs associated with serum AAT likely reflect processes in the liver. One way to infer causality of potentially regulatory SNPs is by testing if they are simultaneously associated with health outcome and gene expression in the relevant tissue [44]. We therefore conducted a look-up in an expression quantitative loci (eQTL) database of lung tissue [45], but could not find any common variant which was significantly associated in *cis* with the transcripts deriving from the *SERPINA1* locus. In a recent study on networks of blood

metabolites, the SNPs rs11628917 and rs1884549 were the most strongly associated blood and liver eQTLs with respect to *SERPINA1* expression [46]. They both lie in the 3′ untranslated region of *SERPINA1*, but were not associated with blood AAT in our GWAS (P = 0.80 and P = 0.21, respectively). Moreover, we could not detect epistasis between those variants and the deleterious coding variants PI S and Z in terms of AAT serum levels. The absence of such an interaction does not point to regulatory function of the common SNPs [47] and argues in favor of tissue-specific heterogeneity. Forth, the role of *SERPINA1* in selected subgroups of persons exhibiting accelerated lung function decline or COPD needs to be considered from a perspective beyond genetic variation, as a recent study investigating epigenetic mechanisms of disease revealed methylation status of the *SERPINA1* gene to be most strongly associated with cross-sectional lung function and COPD [48].

The strength of this study is that it combines the report of a GWAS on AAT serum levels with meta-analyses of the associations of some of the GWAS top variants with lung function. The effects of the underlying functional variants are thoroughly investigated resulting in the hitherto largest meta-

STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.198	-0.004	0.013		
BHS (N=1977)	0.220	-0.013	0.019		
COPENHAGEN (N=6484)	0.185	-0.003	0.013		
KORA_S3 (N=536)	0.185	-0.094	0.038		
KORA_F4 (N=802)	0.196	-0.029	0.033		
LBC36 (N=554)	0.206	0.069	0.039		
LHS (N=542)	0.206	0.024	0.022		
NFBC66 (N=3324)	0.210	0.000	0.005		
SAPALDIA (N=714)	0.170	-0.003	0.038		
SHIP (N=995)	0.210	0.031	0.028		
META_FIXED (N=20153)		-0.001	0.004	0.90	31.4
POPULATION ISOLATES					
KORCULA (N=423)	0.222	0.010	0.040		
NSPHS (N=121)	0.289	-0.107	0.013		
ORCADES (N=288)	0.090	-0.052	0.066		
SPLIT (N=254)	0.222	-0.038	0.052		
VIS (N=537)	0.240	-0.081	0.050		
META_FIXED		-0.091	0.012	<0.0001	56.1
META_RANDOM (N=1623)		-0.061	0.027	0.02	

Change in FEV1 per minor allele

Figure 5. Forest plot of meta-analyzed results for the effect per minor allele of rs4905179 on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2 > 0.5$. Study weights (blue squares) correspond to fixed effect meta-analyses.
doi:10.1371/journal.pgen.1003585.g005

analysis of PI S and Z on FEV1 in ever-smokers. Ascertainment and study design of the many participating studies were sufficiently diverse to informatively address heterogeneity in association of common and rarer variants in the *SERPINA* gene cluster with lung function. The strength of the discovery sample is the population-based study design and the detailed characterization of the participants. Sex, age and smoking are important modifiers of AAT blood levels in the general population [43] and were included in all regression models. More refined smoking variables covering smoking intensity were not included as this information is less complete than smoking status in SAPALDIA and would lower the sample size. By excluding samples with elevated hs-CRP values we avoided the masking of AAT deficiencies due to a chronic or acute inflammation. On the methodological side, conditional analysis is a well-established tool for identifying independent signals within a certain locus [38,49,50]. Furthermore, 1000G imputation was able to point to the causal variant demonstrating its reliability to correctly assign alleles close to the 1% MAF threshold.

The limitations of this investigation include firstly the small sample size of the GWAS discovery arm, resulting in a high susceptibility to false negative findings. We calculated 63% power to detect SNPs with an allele effect of 0.1 g/L AAT serum level (=2.4% of the phenotypic variance) to a genome-wide significance level of $5*10^{-8}$. However, if we define the clinically important

threshold of AAT as the upper limit of intermediate AAT deficiency, which has been recently suggested as 0.92 g/L [20], we have more than 99.9% power to detect such a large-impact variant. Nevertheless, genes that contribute to AAT serum levels with smaller effects than *SERPINA1* were likely to be missed. This could be a reason why neither SNPs in interleukin 6 (IL-6) nor in hepatocyte nuclear factor 1α (HNF-1α)/HNF-4, both important regulators of AAT expression [51], were associated with circulating AAT concentrations. Furthermore, by sequencing only the coding region of *SERPINA1*, rare variants in introns and outside the gene could not be determined. Another potential limitation of our GWAS on AAT serum level is the overrepresentation of asthmatics in the discovery sample. Asthma patients usually show higher levels of inflammatory markers in their lungs. However, we did not find heterogeneity in the effects of the most strongly associated SNPs when comparing asthmatics with non-asthmatics. Moreover, AAT mean values between the discovery and the replication arm were not significantly different, as participants with elevated hs-CRP had been excluded.

In conclusion, our study confirms the *SERPINA1* locus as the major genetic determinant of AAT blood levels. Methodologically, it represents a powerful example how low-frequent variants, separated by several kilobases from the top-ranking GWAS signals, can create purely synthetic associations which do not add to the variance of the respective outcome. In terms of lung function, our data do not support a functional role of

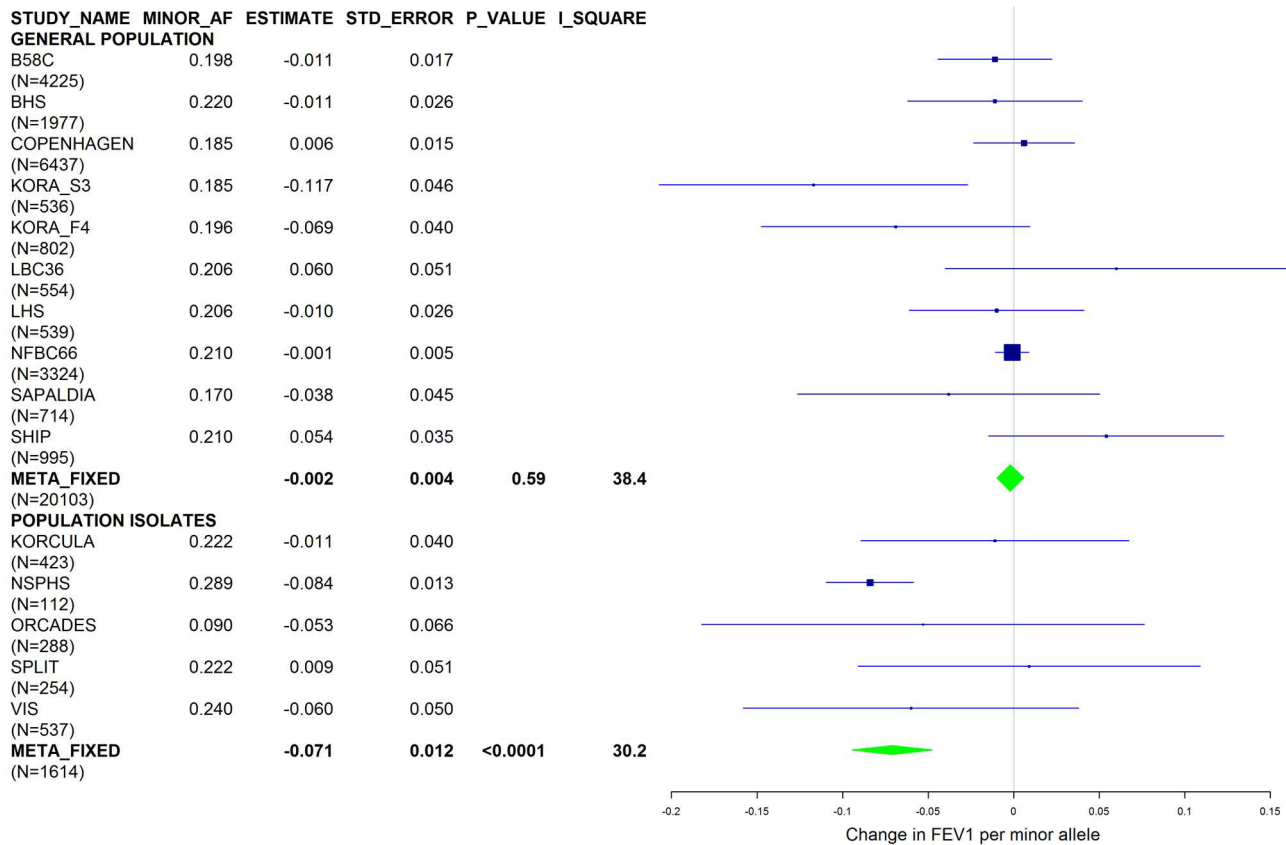| STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.198	-0.011	0.017		
BHS (N=1977)	0.220	-0.011	0.026		
COPENHAGEN (N=6437)	0.185	0.006	0.015		
KORA_S3 (N=536)	0.185	-0.117	0.046		
KORA_F4 (N=802)	0.196	-0.069	0.040		
LBC36 (N=554)	0.206	0.060	0.051		
LHS (N=539)	0.206	-0.010	0.026		
NFBC66 (N=3324)	0.210	-0.001	0.005		
SAPALDIA (N=714)	0.170	-0.038	0.045		
SHIP (N=995)	0.210	0.054	0.035		
META_FIXED (N=20103)		-0.002	0.004	0.59	38.4
POPULATION ISOLATES					
KORCULA (N=423)	0.222	-0.011	0.040		
NSPHS (N=112)	0.289	-0.084	0.013		
ORCADES (N=288)	0.090	-0.053	0.066		
SPLIT (N=254)	0.222	0.009	0.051		
VIS (N=537)	0.240	-0.060	0.050		
META_FIXED (N=1614)		-0.071	0.012	<0.0001	30.2

Figure 6. Forest plot of meta-analyzed results for the effect per minor allele of rs4905179 on FEV1 in ever-smokers, adjusted for sex, age, height, population stratification factors and the presence of PI S and Z alleles. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2 > 0.5$. Study weights (blue squares) correspond to fixed effect meta-analyses.

doi:10.1371/journal.pgen.1003585.g006

any common SNP in the *SERPINA* cluster in the general population.

Materials and Methods

Ethics Statement

SAPALDIA was approved by the Swiss Academy of Medical Sciences, the national ethics committee for clinical research (UREK, Project Approval Number 123/00) and the Cantonal Ethics Committees for each of the eight examination areas (Ethics commissions of the cantons Aargau, Basel, Geneva, Grisons, Ticino, Valais, Vaud and Zurich). Participants were required to give written consent before any part of the health examination was conducted either globally (for all health examinations) or separately for each investigation. For ethics statements of the additional studies contributing to this work, see Table S8.

Study Population

SAPALDIA. In 1991, a random sample of 9651 adults, aged 18–60 years, from eight areas in Switzerland responded to a questionnaire about respiratory health, occupational and lifestyle exposures. 99.0% of them also underwent spirometry testing [52]. Eleven years later, 8047 persons were reassessed and 6058 subjects provided blood samples and consented to

DNA analysis [53]. In the present study, we used a subgroup of the second survey (N = 1640) that underwent genotyping in the context of the GWAS on asthma by the GABRIEL consortium [17]. This sample included all asthmatics (positive answer to the question "Have you ever had asthma?" at either survey) as well as a random sample of non-asthmatic controls. 248 participants were removed due to several reasons, including elevated levels of the inflammatory marker hs-CRP (>10 mg/L, N = 54), leading to a discovery arm of 1392 individuals (Figure S1). The discovery arm contained 548 (39.4%) self-declared asthmatics, whereas there were no self-declared asthmatics in the replication arm. Both the discovery and the replication sample were submitted to a first step of fine mapping of *SERPINA1* resulting in 5569 individuals from whom all the selected SNPs could be successfully determined. In a second step, a subsample with abnormally low AAT measurements additionally underwent *SERPINA1* exon sequencing.

Additional Studies. The populations are briefly described in Table S8.

Phenotype Measurements

AAT serum levels in SAPALDIA were determined by latex-enhanced immunoturbidimetric assays (Roche Diagnostics, on a Roche Cobas Integra analyzer) with interassay coefficients of

STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.152	0.029	0.015		
BHS (N=1977)	0.150	0.015	0.024		
KORA_S3 (N=536)	0.155	0.028	0.040		
KORA_F4 (N=802)	0.117	0.041	0.043		
LBC36 (N=554)	0.158	0.041	0.046		
NFBC66 (N=3324)	0.152	0.010	0.006		
SAPALDIA (N=3037)	0.148	0.003	0.018		
SHIP (N=995)	0.156	-0.049	0.032		
META_FIXED (N=15450)		**0.011**	**0.005**	**0.02**	**0.0**
POPULATION ISOLATES					
KORCULA (N=423)	0.117	0.062	0.052		
NSPHS (N=121)	0.006	0.189	0.028		
ORCADES (N=288)	0.080	-0.040	0.068		
SPLIT (N=254)	0.118	0.023	0.071		
VIS (N=537)	0.175	0.029	0.058		
META_FIXED		**0.114**	**0.021**	**<0.0001**	**76.7**
META_RANDOM (N=1623)		**0.063**	**0.048**	**0.19**	
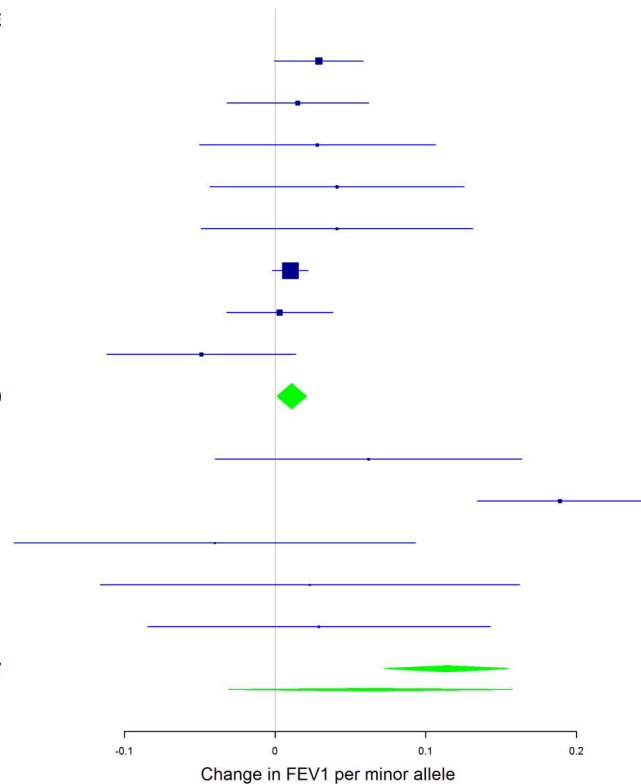
Figure 7. Forest plot of meta-analyzed results for the effect per minor allele of rs3748312 on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2 > 0.5$. Study weights (blue squares) correspond to fixed effect meta-analyses.
doi:10.1371/journal.pgen.1003585.g007

variation below 5% and lower detection rate of 0.21 g/L. Serum concentrations in Copenhagen were measured by immunoturbidimetric assays (Thermo Scientific, on a Thermo Scientific Konelab analyzer) with coefficients of variation below 5% and lower detection rate of 0.10 g/L.

Lung function was measured in all participating studies by spirometry without bronchodilation ([52] and Table S8). In the patient-based studies, in which a lung resection was carried out (Groningen, UBC), lung function measurements were carried out prior to the intervention.

Genotyping

SAPALDIA. Genomic DNA was extracted from blood samples using the Puregene DNA Isolation Kit (Gentra Systems). Genotyping of the GWA-bound subset was performed on the Illumina Human 610quad array. Asthmatic and non-asthmatic samples were tested in random blinded order to avoid systematic array-related artifacts. 567,589 autosomal SNPs were satisfactorily genotyped (mean call rate: 99.7%). 69,892 were excluded from analysis due to violation of Hardy Weinberg Equilibrium (HWE, $P < 10^{-4}$), low call rate (<97%) or MAF<5%.

Genotyping of GWAS finding rs2566347 on chromosome 3 in the SAPALDIA replication arm was carried out using the MassARRAY iPLEX Gold (Sequenom).

The SNPs selected in the first fine mapping step were genotyped by polymerase chain reaction (PCR) with fluorescently labeled

Taq-Man probes (Vic or Fam labels) on a Light Cycler 480 (Roche Diagnostics). All SNPs were in HWE (P>0.01) [20].

Additional Studies. PI S and Z genotypes were determined by PCR in Copenhagen and LHS as previously described [54,55]. The SNP rs4905179 was genotyped in the course of GWAS projects in B58C, BHS, Copenhagen, Korcula, KORA S3, KORA F4, LBC36, LHS, NSPHS, ORCADES, Split, UBC, and Vis.

Imputation

SAPALDIA. We have carried out genome-wide imputation from 60 CEU HapMap2 (release 22, NCBI build 36) reference panels [18] using MACH 1.0.16 [56] resulting in 2,588,592 autosomal HapMap-based SNPs. 2,168,668 SNPs fulfilled the quality criteria, which are as mentioned above for genotyped SNPs and additionally consisted of an imputation-$r^2 > 0.5$.

Further imputation was carried out in the most promising loci using 566 EUR reference haplotypes from the August 2010 release of 1000G on the MACH (pre-phasing) and Minimac-omp programs. SNPs with an imputation-$r^2 > 0.5$ and MAF>0.1% passed the quality check.

Additional Studies. Rs4905197 was imputed based on 1000G reference panels in Groningen, NFBC66, and SHIP (imputation-$r^2 \geq 0.99$). Rs3748312, rs17580 (PI S) and rs28929474 (PI Z) were 1000G imputed in B58C, BHS, Groningen, Korcula, KORA S3, KORA F4, LBC36, NFBC66, NSPHS, ORCADES,

STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.152	0.025	0.016		
BHS (N=1977)	0.150	0.033	0.027		
KORA_S3 (N=536)	0.155	0.049	0.042		
KORA_F4 (N=802)	0.117	0.052	0.045		
LBC36 (N=554)	0.158	0.045	0.048		
NFBC66 (N=3324)	0.152	0.009	0.006		
SAPALDIA (N=3037)	0.148	0.001	0.019		
SHIP (N=995)	0.156	-0.048	0.033		
META_FIXED (N=15450)		**0.011**	**0.005**	**0.02**	**0.9**
POPULATION ISOLATES					
KORCULA (N=423)	0.117	0.069	0.052		
NSPHS (N=112)	0.006	0.157	0.019		
ORCADES (N=288)	0.080	-0.035	0.067		
SPLIT (N=254)	0.118	0.010	0.071		
VIS (N=537)	0.175	0.015	0.058		
META_FIXED		**0.118**	**0.016**	**<0.0001**	**74.3**
META_RANDOM (N=1614)		**0.055**	**0.043**	**0.20**	

Figure 8. Forest plot of meta-analyzed results for the effect per minor allele of rs3748312 on FEV1 in ever-smokers, adjusted for sex, age, height, population stratification factors and the presence of PI S and Z alleles. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2>0.5$. Study weights (blue squares) correspond to fixed effect meta-analyses.
doi:10.1371/journal.pgen.1003585.g008

Split, SHIP, UBC, and Vis (imputation-r^2 0.86–0.98, 0.69–0.83, and 0.82–0.98, respectively).

SERPINA1 SNP Selection for Fine Mapping in SAPALDIA

In an attempt to find AAT modifying *SERPINA1* gene variants acting independently of each other, a multiple strategy to optimally cover the gene was applied. Sequencing of the whole

SERPINA1 gene in 25 unrelated samples from the Italian registry of AAT deficiency which demonstrated extreme phenotypes was used to identify common SNPs not present in HapMap. Extreme phenotypes consisted of 11 samples with AAT>1.60 g/L and hs-CRP <8 mg/L, 3 samples with PI ZZ or PI SZ genotype and AAT<0.20 g/L, 2 samples with PI MZ genotype and AAT<0.60 g/L, as well as 9 non-carriers of PI S or PI Z alleles

Table 5. Minor allele effects on FEV1 of low-frequent and common SNPs in the *SERPINA* gene cluster in ever-smokers undergoing lung resection.

SNP	MAF	Imp-r^2	Allele Effect (L)	P	MAF	Imp-r^2	Allele Effect (L)	P
	Groningen (N = 133)				University of British Columbia, UBC (N = 264)			
rs17580 (PI S)	0.055	0.72	0.42	0.10	0.079	0.76	−0.15	0.29
rs28929474 (PI Z)	0.123	0.98	−0.68	<0.001	0.032	0.98	−0.81	<0.0001
rs4905179	0.299	1.00	−0.22	0.04	0.249	na[a]	−0.23	0.002
rs4905179, adjusted for PI S and Z			−0.07	0.65			−0.08	0.36
rs3748312	0.233	0.95	−0.34	0.008	0.148	0.94	−0.16	0.07
rs3748312, adjusted for PI S and Z			0.16	0.37			0.09	0.35

Abbreviations: FEV1, forced expiratory volume in one second; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Imp-r^2 is an indicator for imputation quality. The analyses were adjusted for age, sex and height.
[a]Rs4905179 was genotyped in UBC.
doi:10.1371/journal.pgen.1003585.t005

with blood levels >0.65 and <1.10 g/L. In these 25 samples, a total of 129 mutations were identified in the *SERPINA1* gene. After removing SNPs which were monomorphic in our data, SNPs deviating from HWE or lying in high LD with an adjacent marker (D′>0.8 and r^2>0.4 according to JLIN [57]), we finally obtained a list of 22 common SNPs (Table 4, selection A). In a second strategy, HapMap CEU data was used to select tagging SNPs (Haploview 3.32) [58], resulting in 8 polymorphisms (selection B). Third, TAMAL [59] was used to identify promising SNPs in the region of the *SERPINA1* gene (selection C). Pairwise LD and the feasibility of designing a corresponding TaqMan assay reduced the number of SNPs to 13. Two established (PI S and Z) and one suggestive (rs8004738 [60]) functional SNPs were added (selection D), resulting in 16 SNPs used in the conditional analysis. Three of them were already part of the SNP array genotyped for the GWAS. The 5 SNPs in coding regions (exons 2–5) were all non-synonymous.

Exon Sequencing in SAPALDIA

We sequenced 410 individuals with abnormally low AAT levels with the Sanger chain-termination method. Different thresholds according to the deficiency genotypes and hs-CRP values were applied to define an abnormally low AAT concentration (PI MM: 1.13 g/L if hs-CRP >8 mg/L and 1.00 g/L if hs-CRP ≤8 mg/L; PI MS: 0.85 g/L; PI MZ: 0.65 g/L) [21]. The cut-off of 1.13 g/L was earlier reported to be the best to differentiate AAT-deficient patients from healthy individuals [61]. Since exon 1 is non-coding, the sequencing procedure was only applied to exons 2 to 5.

Statistical Analysis

AAT serum levels were only marginally skewed to the right, and a log-transformation of these data was omitted since it led to a stronger deviation from normality. Student's t-test was used to compare adjusted mean AAT levels between different subgroups of the sequenced samples. The genome-wide association of 2.17 million quality-controlled SNPs with serum AAT levels was assessed using fixed effects linear regression with ProbABEL [62]. An additive genetic model was applied and the association was adjusted for sex, age, study center, dichotomous current smoking status, as well as population stratification factors. To account for population stratification, we relied on previously inferred ancestry-informative principal components using EIGENSTRAT 2.0 software [63] and HapMap data, as well as additional reference European samples [64]. Cryptic relatedness was detected based on identity-by-state (IBS) analysis. Influence of additional suggestive determinants of AAT, such as hs-CRP, BMI, alcohol intake and passive smoking was assessed in a sensitivity analysis. We also performed genome-wide analysis conditioned on the functionally established PI S and PI Z variants. Bonferroni correction for multiple testing was applied, resulting in $P<5*10^{-8}$ to designate genome-wide significance, taking account of one million independent tests for common variants across the genome. For the SNPs imputed by using 1000G reference samples, we considered a three times lower p-value as adequate as roughly three times more SNPs on chromosome 14 passed an imputation-r^2 threshold of 0.5 (219,471 1000G-derived variants vs. 82,296 HapMap2-derived variants). Applying this to a 2 Mb chromosomal stretch (with approximately 600 HapMap2-derived SNPs) resulted in a significance threshold of roughly $3*10^{-5}$.

For the replicated SNP in the SAPALDIA replication arm, as well as for the lung function analysis, a two-sided p-value of 0.05 was considered significant. We investigated heterogeneity between asthmatics and non-asthmatics in the discovery arm by testing for

a difference between the two effects, using a chi-square test with one degree of freedom.

Replication analysis for AAT in Copenhagen, as well as association analyses of the 16 genotyped *SERPINA1* SNPs in both the SAPALDIA discovery and replication arm, was carried out applying the same statistical model as in the GWAS apart from the adjustment for population stratification factors. Stepwise conditional analyses were conducted by testing each SNP for AAT association after including at each step the most significantly associated SNP in the model. As some of these SNPs turned out to be in unexpectedly high LD, we applied a threshold level for statistical significance of P = 0.005, accounting for approximately ten independent tests [65].

To be as close as possible to the calculations carried out in the original publication [14], multivariate linear regression models for lung function analyses were used adjusted for sex, age, height and population stratification factors (if available).

All the SAPALDIA regression analyses were performed with STATA 12.1 IC.

Further Software

Manhattan, Q-Q and forest plots were created with the help of R 2.15.1 (www.r-project.org). Regional association plots were drawn using LocusZoom [66]. Pairwise LD was calculated for HapMap2 and 1000G CEU data using SNAP [67]. The LD plot was produced with HaploView 4.2 [58]. The effect of non-synonymous SNPs on protein structure was predicted by SIFT [68]. Finally, Quanto 1.2.4 (hydra.usc.edu/gxe/) was used for power calculations for the GWAS.

Supporting Information

Figure S1 SAPALDIA study design for the determination of AAT associated genetic variants.[a] consisting of subjects with abnormally low AAT levels independent of PI S or Z alleles (see Materials and Methods).
(TIF)

Figure S2 Q-Q plot of genome-wide -log(10) p-values for association with AAT serum level.
(TIF)

Figure S3 Forest plot of meta-analyzed results for the effect per minor allele of rs17580 (PI S) on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if I^2>0.5. Study weights (blue squares) correspond to the fixed effect meta-analyses.
(TIF)

Figure S4 Forest plot of meta-analyzed results for the effect per minor allele of rs28929474 (PI Z) on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if I^2>0.5. Study weights (blue squares) correspond to the fixed effect meta-analyses.
(TIF)

Table S1 Characteristics of SAPALDIA follow-up participants belonging to the discovery (N = 1392) and replication arm

(N = 4245), and of participants of the Copenhagen City Heart Study (N = 8273).
(DOC)

Table S2 Characteristics of study populations contributing to the SNP association analyses with FEV1.
(XLS)

Table S3 The top 100 ranking SNPs associated with AAT serum level in SAPALDIA (N = 1392).
(DOC)

Table S4 *SERPINA* regional variants based on 1000 Genomes imputation reaching statistical significance for the association with AAT serum level in SAPALDIA (N = 1392).
(DOC)

Table S5 The top 100 ranking SNPs associated with AAT serum level, conditional on PI S and Z alleles in SAPALDIA (N = 1392).
(DOC)

Table S6 Accuracy of 1000 Genomes based imputation in the *SERPINA1* region in SAPALDIA (N = 1392).
(DOC)

Table S7 Further variants in the *SERPINA1* coding region, present in a SAPALDIA subsample with abnormally low AAT serum levels (N = 410).
(DOC)

Table S8 Descriptions and acknowledgments of individual studies contributing to the SNP association analyses with FEV1.
(XLS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: ML NMPH. Performed the experiments: GAT MI IF MZ MH. Analyzed the data: GAT MI IF AK MO MD SFN ACA DPS VEJ EA JSR AT LML AJS JEH SE. Contributed reagents/materials/analysis tools: AK MO IC FK. Wrote the paper: GAT NMPH. Designed and managed the studies contributing to this project: MI YB KH WT UG OP JFW IR CH AJS IJD BK ER HS JH ALJ TR EWR DPS MRJ IPH MDT MD SFN BGN FK ML NMPH.

References

1. Brantly ML, Wittes JT, Vogelmeier CF, Hubbard RC, Fells GA, et al. (1991) Use of a highly purified alpha 1-antitrypsin standard to establish ranges for the common normal and deficient alpha 1-antitrypsin phenotypes. Chest 100: 703–708.
2. Lomas DA, Evans DL, Finch JT, Carrell RW (1992) The mechanism of Z alpha 1-antitrypsin accumulation in the liver. Nature 357: 605–607.
3. Luisetti M, Seersholm N (2004) Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax 59: 164–169.
4. Lieberman J, Winter B, Sastre A (1986) Alpha 1-antitrypsin Pi-types in 965 COPD patients. Chest 89: 370–373.
5. (2003) American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. Am J Respir Crit Care Med 168: 818–900.
6. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, et al. (2004) Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax 59: 843–849.
7. Dahl M, Hersh CP, Ly NP, Berkey CS, Silverman EK, et al. (2005) The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J 26: 67–76.
8. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, et al. (2010) alpha-Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. Chest 138: 1125–1132.
9. Thun GA, Ferrarotti I, Imboden M, Rochat T, Gerbase M, et al. (2012) SERPINA1 PiZ and PiS Heterozygotes and Lung Function Decline in the SAPALDIA Cohort. PLoS One 7: e42728.
10. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, et al. (2012) Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. Am J Respir Crit Care Med 186: 622–632.
11. Artigas MS, 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.
12. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, et al. (2012) Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol 129: 1218–1228.
13. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, et al. (2011) Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. Am J Respir Crit Care Med 183: 43–49.
14. Obeidat M, Wain LV, Shrine N, Kalsheker N, Artigas MS, et al. (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS One 6: e19382.
15. Law RH, Zhang Q, McGowan S, Buckle AM, Silverman GA, et al. (2006) An overview of the serpin superfamily. Genome Biol 7: 216.
16. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. PLoS Biol 8: e1000294.
17. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363: 1211–1221.
18. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–861.
19. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073.
20. Ferrarotti I, Thun GA, Zorzetto M, Ottaviani S, Imboden M, et al. (2012) Serum levels and genotype distribution of alpha1-antitrypsin in the general population. Thorax 67: 669–674.
21. Zorzetto M, Russi E, Senn O, Imboden M, Ferrarotti I, et al. (2008) SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. Clin Chem 54: 1331–1338.
22. Miyake K, Suzuki H, Oka H, Oda T, Harada S (1979) Distribution of alpha 1-antitrypsin phenotypes in Japanese: description of Pi M subtypes by isoelectric focusing. Jinrui Idengaku Zasshi 24: 55–62.
23. Nukiwa T, Takahashi H, Brantly M, Courtney M, Crystal RG (1987) alpha 1-Antitrypsin nullGranite Falls, a nonexpressing alpha 1-antitrypsin gene associated with a frameshift to stop mutation in a coding exon. J Biol Chem 262: 11999–12004.
24. Graham A, Kalsheker NA, Newton CR, Bamforth FJ, Powell SJ, et al. (1989) Molecular characterisation of three alpha-1-antitrypsin deficiency variants:

proteinase inhibitor (Pi) nullcardiff (Asp256----Val); PiMmalton (Phe51----deletion) and PiI (Arg39----Cys). Hum Genet 84: 55–58.

25. Holmes MD, Brantly ML, Crystal RG (1990) Molecular analysis of the heterogeneity among the P-family of alpha-1-antitrypsin alleles. Am Rev Respir Dis 142: 1185–1192.

26. Okayama H, Brantly M, Holmes M, Crystal RG (1991) Characterization of the molecular basis of the alpha 1-antitrypsin F allele. Am J Hum Genet 48: 1154–1158.

27. Faber JP, Poller W, Weidinger S, Kirchgesser M, Schwaab R, et al. (1994) Identification and DNA sequence analysis of 15 new alpha 1-antitrypsin variants, including two PI*Q0 alleles and one deficient PI*M allele. Am J Hum Genet 55: 1113–1121.

28. Poller W, Merklein F, Schneider-Rasp S, Haack A, Fechner H, et al. (1999) Molecular characterisation of the defective alpha 1-antitrypsin alleles PI Mwurzburg (Pro369Ser), Mheerlen (Pro369Leu), and Q0lisbon (Thr68Ile). Eur J Hum Genet 7: 321–331.

29. Fra AM, Gooptu B, Ferrarotti I, Miranda E, Scabini R, et al. (2012) Three new alpha1-antitrypsin deficiency variants help to define a C-terminal region regulating conformational change and polymerization. PLoS One 7: e38405.

30. Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135–145.

31. Sanna S, Li B, Mulas A, Sidore C, Kang HM, et al. (2011) Fine mapping of five Loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. PLoS Genet 7: e1002198.

32. Oosterveer DM, Versmissen J, Defesche JC, Sivapalaratnam S, Yazdanpanah M, et al. (2013) Low-density lipoprotein receptor mutations generate synthetic genome-wide associations. Eur J Hum Genet 21: 563–566.

33. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, et al. (2010) Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet 42: 1049–1051.

34. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, et al. (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43: 1066–1073.

35. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, et al. (2010) Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42: 684–687.

36. Coassin S, Schweiger M, Kloss-Brandstatter A, Lamina C, Haun M, et al. (2010) Investigation and functional characterization of rare genetic variants in the adipose triglyceride lipase in a large healthy working population. PLoS Genet 6: e1001239.

37. Kronenberg F, Utermann G (2013) Lipoprotein(a): resurrected by genetics. J Intern Med 273: 6–30.

38. Lin JP, Schwaiger JP, Cupples LA, O'Donnell CJ, Zheng G, et al. (2009) Conditional linkage and genome-wide association studies identify UGT1A1 as a major gene for anti-atherogenic serum bilirubin levels–the Framingham Heart Study. Atherosclerosis 206: 228–233.

39. Heid IM, Wagner SA, Gohlke H, Iglseder B, Mueller JC, et al. (2006) Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. Diabetes 55: 375–384.

40. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, et al. (2009) Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41: 47–55.

41. Chappell S, Daly L, Morgan K, Guetta Baranes T, Roca J, et al. (2006) Cryptic haplotypes of SERPINA1 confer susceptibility to chronic obstructive pulmonary disease. Hum Mutat 27: 103–109.

42. Silverman EK, Province MA, Rao DC, Pierce JA, Campbell EJ (1990) A family study of the variability of pulmonary function in alpha 1-antitrypsin deficiency. Quantitative phenotypes. Am Rev Respir Dis 142: 1015–1021.

43. Senn O, Russi EW, Schindler C, Imboden M, von Eckardstein A, et al. (2008) Circulating alpha1-antitrypsin in the general population: determinants and association with lung function. Respir Res 9: 35.

44. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 452: 423–428.

45. 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.

46. Inouye M, Ripatti S, Kettunen J, Lyytikainen LP, Oksala N, et al. (2012) Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet 8: e1002907.

47. Lappalainen T, Montgomery SB, Nica AC, Dermitzakis ET (2011) Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 89: 459–463.

48. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, et al. (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am J Respir Crit Care Med 185: 373–381.

49. 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.

50. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, et al. (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet 43: 1193–1201.

51. Rollini P, Fournier RE (1999) The HNF-4/HNF-1alpha transactivation cascade regulates gene activity and chromatin structure of the human serine protease inhibitor gene cluster at 14q32.1. Proc Natl Acad Sci USA 96: 10308–10313.

52. 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 Praventivmed 42: 67–84.

53. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, et al. (2005) Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991–2003: methods and characterization of participants. Soz Praventivmed 50: 245–263.

54. Dahl M, Tybjaerg-Hansen A, Lange P, Vestbo J, Nordestgaard BG (2002) Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha1-antitrypsin MZ heterozygotes: A longitudinal study of the general population. Ann Intern Med 136: 270–279.

55. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, et al. (2001) Susceptibility genes for rapid decline of lung function in the Lung Health Study. American Journal of Respiratory and Critical Care Medicine 163: 469–473.

56. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34: 816–834.

57. Carter KW, McCaskie PA, Palmer LJ (2006) JLIN: a java based linkage disequilibrium plotter. BMC Bioinformatics 7: 60.

58. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263–265.

59. Hemminger BM, Saelim B, Sullivan PF (2006) TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. Bioinformatics 22: 626–627.

60. Chappell S, Hadzic N, Stockley R, Guetta-Baranes T, Morgan K, et al. (2008) A polymorphism of the alpha1-antitrypsin gene represents a risk factor for liver disease. Hepatology 47: 127–132.

61. Gorrini M, Ferrarotti I, Lupi A, Bosoni T, Mazzola P, et al. (2006) Validation of a rapid, simple method to measure alpha1-antitrypsin in human dried blood spots. Clin Chem 52: 899–901.

62. Aulchenko YS, Struchalin MV, van Duijn CM (2010) ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11: 134.

63. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.

64. Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, et al. (2008) Investigation of the fine structure of European populations with applications to disease association studies. Eur J Hum Genet 16: 1413–1429.

65. Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 95: 221–227.

66. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, et al. (2010) LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26: 2336–2337.

67. 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.

68. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

Table S1. Characteristics of SAPALDIA follow-up participants belonging to the discovery (N=1392) and replication arm (N=4245), and of participants of the Copenhagen City Heart Study (N=8273).

	SAPALDIA discovery	SAPALDIA replication	Copenhagen City Heart Study
Age (mean; SD), years	52.17; 11.18	52.17; 11.40	57.91; 15.12
% women	51.29	50.27	55.12
% current smokers	23.20	25.06 (N=4237)	48.29
% asthmatics[a]	39.37	0 (N=4239)	6.01
FEV1 (mean; SD), mL	3530; 859 (N=1329)	3607; 828 (N=4032)	2728; 1021
% airflow obstructive[b]	25.14 (N=1281)	20.48 (N=3916)	17.42
AAT blood levels (mean; SD), g/L	1.257; 0.200	1.255; 0.199	1.339; 0.276

Abbreviations: AAT, alpha1-antitrypsin; FEV1, forced expiratory volume in one second; FVC, forced vital capacity; SD, standard deviation.
[a] Defined as self-reported at baseline or follow-up.
[b] Defined as pre-bronchodilation FEV1/FVC ratio <0.7.

Table S2. Characteristics of study populations contributing to the SNP association analyses with FEV1.

Study	Sample Size (Ever Smokers)	Women [%]	Age [years], mean; SD	FEV1 [ml], mean; SD	Current smokers [%]	Height [cm], mean; SD	BMI [kg/m2], mean; SD
British 1958 Birth Cohort (B58C)	4225	49.2	45.1; 0.4	3268; 473	33.2	169.6; 6.3	27.3; 4.8
Busselton Health Study (BHS)	1977	46.1	53.5; 15.9	2980; 951	28.0	169.6; 9.2	26.5; 4.2
Copenhagen City Heart Study (Copenhagen)	6437	51.8	58.6; 14.4	2671; 993	62.6	168.6; 9.4	25.6; 4.3
Kooperative Gesundheitsforschung in der Region Augsburg (KORA S3)	536	45.9	44.2; 8.6	3617; 821	45.2	171.3; 9.2	25.5; 3.8
Kooperative Gesundheitsforschung in der Region Augsburg (KORA F4)	802	48.0	51.5; 5.7	3368; 807	36.2	170.8; 9.1	27.6; 5.1
Lothian Birth Cohort 1936 (LBC36)	554	42.2	69.6; 0.8	2332; 705	22.2	167.0; 8.7	28.0; 4.5
Lung Health Study (LHS)	539	37.5	48.6; 6.7	2591; 574	100.0	171.3; 8.6	na
Northern Finland Birth Cohort 1966 (NFBC66)	3324	52.0	31.0; 0.2	3973; 788	61.5	171.5; 9.1	24.8; 4.3
SAPALDIA	3037	44.4	42.0; 10.5	3617; 815	56.0	170.2; 8.6	24.0; 3.6
Study of Health in Pomerania (SHIP)	995	39.0	46.5; 13.8	3380; 871	57.9	171.5; 8.6	28.1; 4.8
CROTIA-Korcula (Korcula)	423	55.1	52.7; 13.1	2892; 748	46.3	170.1; 8.9	27.5; 4.2
Northern Swedish Population Health Study (NSPHS)	121	52.1	51.5; 16.8	2788; 1005	100.0	164.3; 9.8	25.9; 4.3
Orkney Complex Disease Study (ORCADES)	288	43.8	55.7; 15.0	2903; 828	20.8	168.6; 8.5	27.9; 4.8
CROTIA-Split (Split)	254	53.1	49.1; 13.4	3207; 876	48.0	173.7; 9.1	26.9; 4.2
CROTIA-Vis (Vis)	537	43.6	53.1; 14.8	3677; 116	45.5	170.1; 9.5	27.2; 4.4
eQTL-Groningen (Groningen)	133	47.7	57.2; 10.2	2200; 900	23.0	172.1; 10.2	23.9; 3.9
eQTL-University of British Columbia (UBC)	264	44.3	60.7; 13.2	2200; 830	37.9	167.9; 10.4	25.5; 5.3

Table S3. The top 100 ranking SNPs associated with AAT serum level in SAPALDIA (N=1392).

SNP	Chromosome	Position	Gene	Location	Determination	MAF	Imp-r^2	Allele Effect	P
rs2736887	14	93882733		intergenic	imputed	0.185	0.950	0.071	2.48E-13
rs926144	14	93883155		intergenic	imputed	0.186	0.950	0.071	2.72E-13
rs7151526	14	93933389	SERPINA1	5'UTR	imputed	0.065	0.769	0.116	6.78E-13
rs4905179	14	93865245	SERPINA6	5'UTR	genotyped	0.180	1.000	0.068	1.20E-12
rs11621961	14	93839229	SERPINA6	3'UTR	genotyped	0.355	0.945	0.052	1.37E-11
rs17751837	14	93937997	SERPINA1	5'UTR	genotyped	0.097	0.995	0.063	8.56E-08
rs1028580	14	93919635	SERPINA1	intron	imputed	0.154	0.979	0.051	4.87E-07
rs8010121	14	93920367	SERPINA1	intron	genotyped	0.155	0.999	0.049	6.64E-07
rs3748312	14	93924017	SERPINA1	intron	imputed	0.148	0.846	0.053	9.84E-07
rs17752593	14	94007781	SERPINA9	intron	genotyped	0.129	0.997	0.053	1.59E-06
rs2566347	3	159974071		intergenic	imputed	0.192	0.998	0.044	1.87E-06
rs4703798	5	79448084	SERINC5	intron	genotyped	0.386	0.999	0.035	2.73E-06
rs11160184	14	94007744	SERPINA9	intron	imputed	0.144	0.933	0.049	4.42E-06
rs1430414	3	159987697	MFSD1	5'UTR	imputed	0.137	0.984	0.048	4.44E-06
rs1560417	3	159972476		intergenic	imputed	0.200	0.998	0.041	5.06E-06
rs1560418	3	159972335		intergenic	genotyped	0.200	1.000	0.041	5.07E-06
rs6761989	3	159983253		intergenic	imputed	0.137	0.993	0.047	5.44E-06
rs17643917	3	159968433		intergenic	imputed	0.137	1.000	0.047	5.99E-06
rs17643860	3	159967954		intergenic	imputed	0.137	1.000	0.047	6.02E-06
rs17700475	3	159967627		intergenic	genotyped	0.137	1.000	0.047	6.04E-06
rs6580353	5	139390871	NRG2	intron	genotyped	0.248	0.948	0.039	6.63E-06
rs6893779	5	129825798		intergenic	imputed	0.188	0.980	0.041	7.63E-06
rs6595951	5	129874072		intergenic	genotyped	0.182	0.999	0.042	8.17E-06
rs12089980	1	150698555		intergenic	genotyped	0.091	1.000	0.054	9.47E-06
rs7731657	5	129971218		intergenic	imputed	0.170	0.942	0.044	1.01E-05

rs4836515	5	129965471		intergenic	imputed	0.170	0.951	0.044	1.03E-05
rs10069896	5	129938264		intergenic	imputed	0.171	0.971	0.043	1.04E-05
rs10035791	5	129919660		intergenic	genotyped	0.172	1.000	0.042	1.05E-05
rs10039487	5	129980823		intergenic	imputed	0.169	0.941	0.044	1.13E-05
rs10276467	7	2701239	*AMZ1*	intron	genotyped	0.089	0.987	0.056	1.18E-05
rs6874868	5	129825862		intergenic	imputed	0.185	0.989	0.041	1.18E-05
rs12516876	5	129818959		intergenic	imputed	0.185	0.989	0.041	1.28E-05
rs7714333	5	6370260	*FLJ33360*	intron	genotyped	0.459	0.972	0.032	1.32E-05
rs16882595	8	112874500		intergenic	imputed	0.243	0.987	0.038	1.33E-05
rs12479315	2	56960739		intergenic	imputed	0.491	0.963	0.032	1.38E-05
rs2216407	2	56963650		intergenic	imputed	0.491	0.969	0.032	1.39E-05
rs12081541	1	150707990		intergenic	imputed	0.091	0.904	0.056	1.43E-05
rs16861952	3	150735514	*WWTR1*	intron	imputed	0.074	0.981	0.060	1.45E-05
rs7141205	14	93838612	*SERPINA6*	3'UTR	genotyped	0.179	0.954	0.043	1.49E-05
rs13181256	5	129806017		intergenic	imputed	0.184	0.980	0.041	1.53E-05
rs4150667	11	18339487	*GTF2H1*	intron	genotyped	0.310	0.999	0.034	1.62E-05
rs4263026	18	71438977		intergenic	imputed	0.198	0.983	0.039	1.72E-05
rs3863076	3	159969394		intergenic	genotyped	0.145	1.000	0.044	1.73E-05
rs12497958	3	160010142	*MFSD1*	intron	imputed	0.109	1.000	0.050	1.82E-05
rs6441226	3	160008917	*MFSD1*	intron	genotyped	0.109	1.000	0.050	1.82E-05
rs256435	5	79441982	*SERINC5*	3'UTR	imputed	0.438	0.789	0.035	1.84E-05
rs7624420	3	159993368	*MFSD1*	5'UTR	imputed	0.102	0.990	0.052	1.85E-05
rs13184933	5	129873782		intergenic	imputed	0.203	0.935	0.039	1.86E-05
rs2041342	7	2696291	*AMZ1*	intron	imputed	0.093	0.924	0.055	1.94E-05
rs9348171	6	167009010	*RPS6KA2*	intron	imputed	0.207	0.916	0.038	2.18E-05
rs12462442	19	18094195	*MAST3*	intron	imputed	0.142	1.000	0.043	2.30E-05
rs11086090	19	18093754	*MAST3*	intron	genotyped	0.142	1.000	0.043	2.30E-05
rs756322	10	72414636		intergenic	imputed	0.089	0.680	0.065	2.51E-05
rs6420430	16	80703651	*HSD17B2*	3'UTR	imputed	0.094	0.961	0.053	2.59E-05
rs1251581	1	76198013	*ASB17*	5'UTR	imputed	0.377	0.956	0.032	2.78E-05

rs10771459	12	29026497		intergenic	imputed	0.436	0.994	0.030	2.80E-05
rs10936162	3	160018817	*MFSD1*	intron	imputed	0.109	0.973	0.050	2.83E-05
rs2399702	15	95849200		intergenic	imputed	0.078	0.638	0.070	2.86E-05
rs12611216	19	18095005	*MAST3*	intron	imputed	0.145	0.984	0.042	2.89E-05
rs8104096	19	2561822	*GNG7*	intron	genotyped	0.253	0.979	0.035	2.89E-05
rs1689271	1	76165450	*ASB17*	intron	imputed	0.365	0.948	0.032	3.05E-05
rs12372252	12	29025493		intergenic	imputed	0.436	0.995	0.030	3.15E-05
rs12605699	18	71433743		intergenic	imputed	0.190	0.857	0.041	3.23E-05
rs17290460	13	73463470	*KLF12*	intron	imputed	0.096	0.927	0.052	3.35E-05
rs2281143	6	167015841	*RPS6KA2*	intron	imputed	0.222	0.986	0.035	3.45E-05
rs4835994	5	129749756		intergenic	imputed	0.173	0.972	0.040	3.46E-05
rs1658737	1	76240227		intergenic	imputed	0.371	0.991	0.031	3.47E-05
rs1251531	1	76241922		intergenic	genotyped	0.371	1.000	0.031	3.48E-05
rs10484524	6	167016267	*RPS6KA2*	intron	genotyped	0.224	0.999	0.035	3.55E-05
rs2623960	6	152935134	*SYNE1*	intron	genotyped	0.232	0.999	0.036	3.57E-05
rs1251551	1	76236265		intergenic	imputed	0.370	0.985	0.031	3.58E-05
rs1251550	1	76234909		intergenic	imputed	0.370	0.989	0.031	3.69E-05
rs17218496	13	73468873	*KLF12*	intron	imputed	0.099	0.950	0.051	3.72E-05
rs10145569	14	70722923	*c14orf56*	3'UTR	imputed	0.400	0.972	0.031	3.76E-05
rs740692	19	18095567	*MAST3*	intron	imputed	0.147	0.980	0.041	3.81E-05
rs6458184	6	40873668		intergenic	imputed	0.271	0.992	0.034	3.86E-05
rs17290439	13	73462261	*KLF12*	intron	imputed	0.098	0.911	0.052	3.92E-05
rs1796812	1	76174419	*ASB17*	5'UTR	imputed	0.356	0.989	0.031	3.94E-05
rs4695888	4	174998288		intergenic	genotyped	0.500	0.998	0.029	3.97E-05
rs10843288	12	29023614		intergenic	imputed	0.437	0.996	0.030	4.03E-05
rs6930345	6	40873538		intergenic	genotyped	0.271	0.999	0.034	4.20E-05
rs7299873	12	29029495		intergenic	imputed	0.440	0.989	0.030	4.28E-05
rs1741547	4	3491737	*LRPAP1*	intron	imputed	0.300	0.949	0.033	4.32E-05
rs6574008	14	70722644	*c14orf56*	3'UTR	imputed	0.402	0.964	0.031	4.39E-05
rs10771458	12	29023602		intergenic	imputed	0.437	0.997	0.030	4.40E-05

rs1794444	4	3487530	*LRPAP1*	intron	imputed	0.300	0.949	0.033	4.48E-05
rs6564966	16	80700610	*HSD17B2*	3'UTR	imputed	0.082	0.906	0.056	4.65E-05
rs2184658	1	219119080	*HLX*	5'UTR	imputed	0.179	0.936	0.039	4.66E-05
rs11049978	12	29030367		intergenic	imputed	0.440	0.992	0.030	4.70E-05
rs1850629	6	118200284		intergenic	imputed	0.092	0.744	0.059	4.73E-05
rs11049971	12	29021467		intergenic	genotyped	0.438	0.999	0.030	4.80E-05
rs740693	19	18095639	*MAST3*	intron	imputed	0.148	0.974	0.040	4.88E-05
rs6789	4	3484493	*LRPAP1*	3'UTR	imputed	0.301	0.952	0.033	4.98E-05
rs1359422	13	73480470	*KLF12*	intron	genotyped	0.098	0.999	0.049	5.03E-05
rs1607369	6	118187203	*NUS1*	3'UTR	imputed	0.091	0.735	0.059	5.20E-05
rs973229	5	129775011		intergenic	imputed	0.171	0.983	0.039	5.26E-05
rs3738182	1	219124285	*HLX*	exon	imputed	0.178	0.933	0.039	5.40E-05
rs7303799	12	29033635		intergenic	imputed	0.425	0.957	0.030	5.45E-05
rs2189378	5	129763506		intergenic	imputed	0.169	0.980	0.039	5.46E-05
rs1489078	12	29032680		intergenic	genotyped	0.441	1.000	0.029	5.77E-05

Abbreviations: AAT, alpha1-antitrypsin; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

Imp-r^2 is an indicator of imputation quality. SNPs with MAF <0.05 or imp-r^2 <0.5 were excluded.

Chromosomal position is based on reference panel, NCBI build 36.3. Allele Effects are shown in absolute numbers.

Table S4. *SERPINA* regional variants based on 1000 Genomes imputation reaching statistical significance for the association with AAT serum level in SAPALDIA (N=1392).

SNP	Position	Gene	Location	MAF	Imp-r^2	Allele Effect	P	GWAS inclusion
rs28929474	94844947	*SERPINA1*	exon	0.008	0.684	0.620	4.61E-43	no
rs112458284	94672731	*PPP4R4*	intron	0.028	0.578	0.363	5.35E-32	no
rs149837463	94632757		intergenic	0.022	0.529	0.388	1.09E-26	no
rs111974986	94533735	*DDX24*	intron	0.017	0.641	0.332	8.33E-20	no
rs7151526	94863636	*SERPINA1*	5'UTR	0.057	0.702	0.144	5.88E-15	yes
rs61980636	94784618	*SERPINA6*	intron	0.181	0.977	0.070	2.70E-13	no, but in high LD
rs2736887	94812980		intergenic	0.195	0.980	0.068	4.91E-13	yes
rs926144	94813402		intergenic	0.195	0.980	0.068	5.05E-13	yes
rs965344	94818078		intergenic	0.195	0.980	0.068	6.62E-13	no, but in high LD
rs4905179	94795492	*SERPINA6*	5'UTR	0.184	1.000	0.068	6.90E-13	yes
rs61280460	94796184	*SERPINA6*	5'UTR	0.184	1.000	0.068	6.98E-13	no, but in high LD
rs7149605	94854041	*SERPINA1*	intron	0.095	0.853	0.084	5.14E-11	no, but in high LD
rs11621961	94769476	*SERPINA6*	3'UTR	0.353	0.986	0.048	9.52E-11	yes
rs75416602	94554409	*IFI27L1*	intron	0.041	0.897	0.122	3.12E-09	no
rs79811936	94510275	*OTUB2*	intron	0.040	0.824	0.122	1.04E-08	no, but in high LD
rs74712407	94506729	*OTUB2*	intron	0.040	0.823	0.122	1.16E-08	no, but in high LD
rs55862705	94866482	*SERPINA1*	5'UTR	0.093	0.949	0.070	1.49E-08	no, but in high LD
rs17751837	94868244	*SERPINA1*	5'UTR	0.099	0.998	0.063	6.54E-08	yes
rs145730801	94768196	*SERPINA6*	3'UTR	0.052	0.608	0.112	7.01E-08	no
rs35306951	94873695		intergenic	0.098	0.971	0.064	9.35E-08	no, but in high LD
rs8015929	94763787		intergenic	0.370	0.635	0.048	1.78E-07	no
rs3748312	94854264	*SERPINA1*	intron	0.161	0.864	0.055	2.09E-07	yes
rs7146221	94769081	*SERPINA6*	3'UTR	0.443	0.902	0.039	3.57E-07	no
rs72692809	94904542	*SERPINA11*	3'UTR	0.104	0.583	0.076	3.64E-07	no

rs55683719	94936222	*SERPINA9*	intron	0.136	0.791	0.058	4.42E-07	no, but in high LD
rs1028580	94849882	*SERPINA1*	intron	0.163	0.998	0.050	5.35E-07	yes
rs8010121	94850614	*SERPINA1*	intron	0.163	1.000	0.049	5.90E-07	yes
rs72704312	94733695	*PPP4R4*	intron	0.072	0.541	0.092	9.09E-07	no
rs61976073	95050970	*SERPINA5*	intron	0.075	0.861	0.067	9.95E-07	no
rs17752593	94938028	*SERPINA9*	intron	0.123	0.981	0.053	1.59E-06	yes
rs11160167	94763621		intergenic	0.495	0.748	0.039	2.29E-06	no
rs56149519	94554061	*IFI27L1*	intron	0.067	0.803	0.077	2.34E-06	no, but in high LD
rs11160184	94937991	*SERPINA9*	intron	0.150	0.903	0.048	5.29E-06	yes
rs11622665	94774995	*SERPINA6*	intron	0.160	0.988	0.045	5.98E-06	yes
rs12892767	94767211	*SERPINA6*	3'UTR	0.444	0.831	0.035	9.58E-06	no
rs61976081	95066611		intergenic	0.084	0.693	0.064	1.55E-05	no

Abbreviations: AAT, alpha1-antitrypsin; LD, linkage disequilibrium; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

Imp-r^2 is an indicator of imputation quality. SNPs with MAF <0.001 or imp-r^2 <0.5 were excluded. Presented are results with P<3*10(-5) within a 2 Mb region from 93,857 kb to 95,857 kb on chromosome 14.

Chromosomal position is based on panel GRCh37.p5, NCBI build 37.3. Allele Effects are shown in absolute numbers. High LD is defined as a LD-r^2 >0.8 using SNAP software.

Table S5. The top 100 ranking SNPs associated with AAT serum level, conditional on PI S and Z alleles in SAPALDIA (N=1392).

SNP	Chromosome	Position	Gene	Location	Determination	MAF	Imp-r^2	Allele Effect	P
rs2566347	3	159974071		intergenic	imputed	0.192	0.998	0.043	7.88E-08
rs1560417	3	159972476		intergenic	imputed	0.200	0.998	0.042	1.11E-07
rs1560418	3	159972335		intergenic	genotyped	0.200	1.000	0.042	1.11E-07
rs1430414	3	159987697	MFSD1	5'UTR	imputed	0.137	0.984	0.045	9.26E-07
rs6761989	3	159983253		intergenic	imputed	0.137	0.993	0.044	1.14E-06
rs17643917	3	159968433		intergenic	imputed	0.137	1.000	0.044	1.23E-06
rs17643860	3	159967954		intergenic	imputed	0.137	1.000	0.044	1.24E-06
rs17700475	3	159967627		intergenic	genotyped	0.137	1.000	0.044	1.25E-06
rs3863076	3	159969394		intergenic	genotyped	0.145	1.000	0.042	1.69E-06
rs2206593	1	184909052	PTGS2	3'UTR	genotyped	0.065	0.956	0.060	4.60E-06
rs12462442	19	18094195	MAST3	intron	imputed	0.142	1.000	0.040	5.74E-06
rs11086090	19	18093754	MAST3	intron	genotyped	0.142	1.000	0.040	5.74E-06
rs9541793	13	68628588		intergenic	imputed	0.258	0.990	0.032	7.37E-06
rs1446378	13	68626607		intergenic	imputed	0.258	0.994	0.032	7.45E-06
rs4703798	5	79448084	SERINC5	intron	genotyped	0.386	0.999	0.029	7.50E-06
rs1446379	13	68626323		intergenic	imputed	0.258	0.995	0.032	7.61E-06
rs13400830	2	12972015		intergenic	genotyped	0.083	0.989	0.051	7.83E-06
rs1023818	13	68625043		intergenic	genotyped	0.258	1.000	0.031	8.04E-06
rs16861952	3	150735514	WWTR1	intron	imputed	0.074	0.981	0.054	8.28E-06
rs16859504	2	12972809		intergenic	imputed	0.081	0.926	0.053	8.38E-06
rs12611216	19	18095005	MAST3	intron	imputed	0.145	0.984	0.039	1.01E-05
rs6682093	1	179309512	MR1	3'UTR	imputed	0.362	0.983	0.029	1.03E-05
rs1119065	1	184924695	PTGS2	5'UTR	imputed	0.066	0.810	0.062	1.03E-05
rs1013665	1	184814763		intergenic	imputed	0.177	0.809	0.040	1.05E-05
rs7514491	1	184865641		intergenic	genotyped	0.173	0.998	0.036	1.19E-05

rs9599423	13	68623359		intergenic	imputed	0.261	0.989	0.031	1.31E-05
rs6662507	1	179322783	*IER5*	5'UTR	imputed	0.346	0.955	0.029	1.42E-05
rs9511045	13	23564937	*SPATA13*	5'UTR	imputed	0.112	0.931	0.044	1.44E-05
rs7300338	12	4010161	*LOC100507511*	3'UTR	imputed	0.384	0.865	0.030	1.61E-05
rs9541789	13	68620213		intergenic	imputed	0.262	0.985	0.030	1.63E-05
rs3856068	1	179312991	*MR1*	3'UTR	imputed	0.365	0.982	0.028	1.67E-05
rs9529519	13	68620322		intergenic	imputed	0.262	0.986	0.030	1.68E-05
rs740692	19	18095567	*MAST3*	intron	imputed	0.147	0.980	0.037	1.74E-05
rs2275470	1	179332430	*IER5*	3'UTR	imputed	0.344	0.924	0.030	1.79E-05
rs12479315	2	56960739		intergenic	imputed	0.491	0.963	0.027	1.86E-05
rs2216407	2	56963650		intergenic	imputed	0.491	0.969	0.027	1.92E-05
rs2304408	3	159803574	*MLF1*	intron	imputed	0.229	0.972	0.032	1.99E-05
rs7624771	3	159818103	*MLF1*	3'UTR	imputed	0.228	0.982	0.032	2.28E-05
rs11923661	3	71284796	*FOXP1*	intron	genotyped	0.446	0.999	0.027	2.34E-05
rs453176	21	45407444	*ADARB1*	intron	imputed	0.053	0.932	0.063	2.37E-05
rs407550	21	45408504	*ADARB1*	intron	imputed	0.053	0.932	0.063	2.37E-05
rs17699324	3	159907255	*RARRES1*	intron	genotyped	0.225	0.998	0.032	2.40E-05
rs17699103	3	159902483	*RARRES1*	intron	imputed	0.225	0.976	0.032	2.41E-05
rs7610009	3	159903063	*RARRES1*	intron	imputed	0.225	0.978	0.032	2.41E-05
rs2745557	1	184915844	*PTGS2*	intron	imputed	0.179	0.882	0.037	2.43E-05
rs3867391	3	159898534	*RARRES1*	intron	imputed	0.226	0.970	0.032	2.44E-05
rs17698754	3	159894398	*GFM1*	3'UTR	imputed	0.226	0.968	0.032	2.46E-05
rs9541803	13	68639250		intergenic	imputed	0.280	0.988	0.029	2.46E-05
rs17004749	21	45418411	*ADARB1*	intron	imputed	0.052	0.941	0.063	2.54E-05
rs6776901	3	159843286	*GFM1*	5'UTR	imputed	0.228	0.991	0.032	2.55E-05
rs6777231	3	159843638	*GFM1*	5'UTR	imputed	0.228	0.989	0.032	2.56E-05
rs17698494	3	159888567	*GFM1*	intron	imputed	0.227	0.969	0.032	2.57E-05
rs740693	19	18095639	*MAST3*	intron	imputed	0.148	0.974	0.037	2.59E-05
rs17697458	3	159859048	*GFM1*	intron	imputed	0.227	0.970	0.032	2.62E-05
rs1522178	3	71286953	*FOXP1*	intron	imputed	0.449	0.972	0.027	2.63E-05

rs7279483	21	45421702	ADARB1	intron	imputed	0.052	0.945	0.063	2.65E-05
rs17630607	3	159872554	GFM1	intron	imputed	0.227	0.974	0.032	2.67E-05
rs4338667	13	67094362		intergenic	genotyped	0.184	0.998	0.034	2.71E-05
rs6971526	7	66998462		intergenic	genotyped	0.479	0.999	0.026	2.74E-05
rs480075	1	184722959	PDC	5'UTR	imputed	0.179	0.693	0.041	2.76E-05
rs9390650	6	149363751	UST	intron	genotyped	0.409	1.000	0.027	2.78E-05
rs9541790	13	68621026		intergenic	genotyped	0.283	0.999	0.028	2.90E-05
rs497285	1	184721052	PDC	5'UTR	imputed	0.180	0.693	0.041	2.94E-05
rs2745559	1	184918625	PTGS2	5'UTR	imputed	0.179	0.862	0.037	3.03E-05
rs1983870	13	23568163	SPATA13	5'UTR	genotyped	0.105	0.992	0.043	3.04E-05
rs7770997	6	149373032	UST	intron	imputed	0.411	0.987	0.027	3.16E-05
rs10172921	2	12960321		intergenic	imputed	0.093	0.956	0.045	3.30E-05
rs9484838	6	144385865	PLAGL1	intron	imputed	0.132	0.942	0.039	3.34E-05
rs10182783	2	12939134		intergenic	imputed	0.124	0.837	0.043	3.37E-05
rs4835994	5	129749756		intergenic	imputed	0.173	0.972	0.035	3.38E-05
rs7653672	3	71284144	FOXP1	intron	imputed	0.448	0.963	0.027	3.45E-05
rs4263026	18	71438977		intergenic	imputed	0.198	0.983	0.033	3.47E-05
rs11099924	4	155070285		intergenic	imputed	0.405	0.920	0.028	3.50E-05
rs13388631	2	12960274		intergenic	imputed	0.093	0.961	0.045	3.56E-05
rs11593082	10	8080437	TAF3	intron	genotyped	0.077	0.997	0.048	3.69E-05
rs447590	13	112495212	ATP11A	intron	imputed	0.427	0.902	0.028	3.75E-05
rs13395897	2	77520241	LRRTM4	intron	imputed	0.078	0.689	0.057	3.76E-05
rs13430556	2	12960160		intergenic	imputed	0.093	0.968	0.045	3.82E-05
rs4411966	4	155074041		intergenic	imputed	0.404	0.917	0.028	3.90E-05
rs2688352	8	3816166	CSMD1	intron	genotyped	0.188	0.993	0.033	3.93E-05
rs12489730	3	159940608	RARRES1	5'UTR	imputed	0.218	0.983	0.031	3.94E-05
rs2460535	10	43141386	RASGEF1A	5'UTR	imputed	0.101	0.899	0.047	3.95E-05
rs2399702	15	95849200		intergenic	imputed	0.078	0.638	0.060	3.95E-05
rs4798376	18	5594240	EPB41L3	5'UTR	genotyped	0.254	0.992	0.030	4.03E-05
rs10145569	14	70722923	c14orf56	3'UTR	imputed	0.400	0.972	0.027	4.03E-05

SNP	Chr	Position	Gene	Region	Type	MAF	Imp-r²	Effect	P-value
rs13279316	8	112676764		intergenic	genotyped	0.260	0.998	0.029	4.09E-05
rs12609740	19	18101298	MAST3	intron	imputed	0.149	0.968	0.036	4.16E-05
rs13390247	2	12963014		intergenic	imputed	0.090	0.961	0.045	4.20E-05
rs12698626	7	66991554		intergenic	imputed	0.476	0.915	0.026	4.20E-05
rs2129818	12	16834202		intergenic	genotyped	0.399	1.000	0.026	4.28E-05
rs7624420	3	159993368	MFSD1	5'UTR	imputed	0.102	0.990	0.043	4.43E-05
rs6574008	14	70722644	c14orf56	3'UTR	imputed	0.402	0.964	0.027	4.51E-05
rs1468488	22	15970744	IL17RA	3'UTR	imputed	0.291	0.922	0.029	4.52E-05
rs6728142	2	12958541		intergenic	imputed	0.092	0.982	0.044	4.54E-05
rs6930963	6	149396494	UST	intron	imputed	0.410	0.954	0.027	4.55E-05
rs6893779	5	129825798		intergenic	imputed	0.188	0.980	0.033	4.63E-05
rs8002029	13	23564535	SPATA13	5'UTR	genotyped	0.081	0.998	0.046	4.65E-05
rs4622983	4	155074124		intergenic	imputed	0.405	0.910	0.027	4.71E-05
rs1386819	12	16847868		intergenic	imputed	0.398	0.967	0.026	4.72E-05
rs7984719	13	23564571	SPATA13	5'UTR	imputed	0.081	0.989	0.046	4.73E-05

Abbreviations: AAT, alpha1-antitrypsin; MAF, minor allele frequency; SNP, single nucleotide polymorphism.
Imp-r^2 is an indicator of imputation quality. SNPs with MAF <0.05 or imp-r^2 <0.5 were excluded.
Chromosomal position is based on reference panel, NCBI build 36.3. Allele Effects are shown in absolute numbers.

Table S6. Accuracy of 1000 Genomes based imputation in the *SERPINA1* region in SAPALDIA (N=1392).

SNP	Location	MAF (genotyped)	MAF (imputed)
rs2896268	5'UTR	0.491	0.488
rs1956707	5'UTR	0.036	0.037
rs8004738	exon 1	0.489	0.478
rs1570142	intron 1	0.488	0.478
rs3748312	intron 1	0.158	0.161
rs3748316	intron 1	0.178	genotyped, Illumina 610quad
rs3748317	intron 1	0.162	0.159
rs1980617	intron 1	0.393	0.385
rs1980618	intron 1	0.388	0.385
rs2753935	intron 1	0.428	0.440
rs2144831	intron 1	0.238	0.238
rs709932	exon 2	0.174	genotyped, Illumina 610quad
rs6647	exon 3	0.208	0.220
rs17580 (S)	exon 3	0.040	0.057
rs28929474 (Z)	exon 5	0.013	0.008
rs1303	exon 5	0.258	genotyped, Illumina 610quad

Abbreviations: GWAS, genome-wide association study; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

Table S7. Further variants in the *SERPINA1* coding region, present in a SAPALDIA subsample with abnormally low AAT serum levels (N=410).

Variant, name	N[a]	Location	Position	Function	Description	Computational prediction [68]
rs28931570, I	9	exon 2	93919141	SNP, non-synonymous	deficiency variant [24]	damaging
ΔF52, Mmalton	3	exon 2	93919100-02	codon deletion	deficiency variant [24]	n/a
rs111850950, M6passau	5	exon 2	93919078	SNP, non-synonymous	neutral variant [27]	damaging
Q105P	1	exon 2	93918942	SNP, non-synonymous	putative deficiency variant [21]	damaging
rs20546, M3riedenburg	17	exon 2	93918904	SNP, synonymous	neutral variant [27]	n/a
rs112030253, V	1	exon 2	93918814	SNP, non-synonymous	neutral variant [27]	tolerated
A153A	1	exon 2	93918797	SNP, synonymous	novel	n/a
ΔY160, Q0granitefalls	1	exon 2	93918776	nucleotide deletion	null variant [23]	n/a
rs28929470, F	3	exon 3	93917139	SNP, non-synonymous	neutral variant [26]	damaging
rs28929472, Plowell	9	exon 3	93917039	SNP, non-synonymous	deficiency variant [25]	damaging
rs1049800, Psaint albans	1	exon 3	93917038	SNP, synonymous	neutral variant [25]	n/a
K259I	6	exon 3	93917030	SNP, non-synonymous	deficiency variant [29]	damaging
T268I	1	exon 3	93917003	SNP, non-synonymous	putative deficiency variant [21]	damaging
D270N[b]	1	exon 3	93916998	SNP, non-synonymous	putative deficiency variant [21]	damaging
rs200945035, Etokyo	1	exon 5	93914721	SNP, non-synonymous	neutral variant [22]	damaging
rs61761869, Mwurzburg	5	exon 5	93914619	SNP, non-synonymous	deficiency variant [28]	damaging

Abbreviations: AAT, alpha1-antitrypsin; MAF, minor allele frequency; SNP, single nucleotide polymorphism; n/a, not applicable.
Chromosomal position is based on reference panel, NCBI build 36.3.
[a] Number of participants carrying the variant heterozygously.
[b] This individual also had the F-variant (rs28929470).

Table S8. Descriptions and acknowledgments of individual studies contributing to the SNP association analyses with FEV1 .

Study	General Description	Acknowledgments
British 1958 Birth Cohort (B58C)	The British 1958 Birth Cohort is an ongoing follow-up of all persons born in England, Scotland and Wales during one week in 1958. At the age of 44-45 years, the cohort were followed up with a biomedical examination, including spirometry as described in more detail elsewhere [Strachan DP, et al. (2007). Int J Epidemiol 36: 522-531] and blood sampling, from which a DNA collection was established as a nationally representative reference panel (http://www.b58cgene.sgul.ac.uk/). Non-overlapping subsets of the DNA collection were genotyped by the Wellcome Trust Case-Control Consortium (WTCCC), the Type 1 Diabetes Genetics Consortium (T1DGC) and the GABRIEL consortium. Genotyping by the WTCCC used both the Affymetrix 500K array and the Illumina 550K (version 1) array. Since the T1DGC used the Illumina 550K (version 3) array and GABRIEL used the Illumina 610K array, a combined dataset was created of SNPs in common across these three panels. SNPs were excluded from subsequent imputation if they had MAF<1%, call-rate<95%, HWE p-value<0.0001, or differences in allele frequency across the three deposits (p<0.0001 on pairwise comparisons). Pre-imputation phasing was performed using MACH. Imputations against the March 2012 release of 1000-genomes all-ethnicities reference haplotypes were performed using Minimac. Associations of imputed allele dosages with lung function, as measured at the 44-45-year follow-up, were analyzed using ProbABEL.	We acknowledge use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council and Wellcome Trust.
Busselton Health Study (BHS)	The Busselton Health Study (BHS) is a longitudinal survey of the town of Busselton in the south-western region of Western Australia that began in 1966. In 1994/1995 a cross-sectional community follow-up study was undertaken where blood was taken for DNA extraction. A sample of 1,168 European-ancestry individuals were genotyped using the Illumina 610-Quad BeadChip (BHS1), and subsequent genotyping was carried out on an independent group of 3,038 European-ancestry individuals (BHS2).	The Busselton Health Study (BHS) acknowledges the generous support for the 1994/1995 follow-up study from Healthway, Western Australia, and the numerous Busselton community volunteers who assisted with data collection and the study participants from the Shire of Busselton. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia.

| Kooperative Gesundheitsforschung in der Region Augsburg (KORA S3, KORA F4) | The KORA studies (Cooperative Health Research in the Region of Augsburg) are a series of independent population based studies from the general population living in the region of Augsburg, Southern Germany [Holle R, et al. (2005). Gesundheitswesen 67 Suppl 1: S19-S25; Wichmann H-E, et al. (2005). Gesundheitswesen 67 Suppl 1: S26-S30]. The KORA S3 study including 4,856 individuals was conducted in 1994/95. Spirometry was measured in 1997/98 for all participants younger than 60 years who did not smoke or use inhalers one hour before the test. All spirometric tests were performed strictly adhering to the ECRHS protocol. Tests were accounted valid if at least two technically satisfactory manoeuvres could be obtained throughout a maximum of nine trials. FEV1 was defined as the maximum value within all valid manoeuvers. Genotyping was performed on the Illumina Omni 2.5 and the Illumina Omni Express platform. Genotypes were called with Genome Studio and annotated to NCBI build 37. We excluded individuals with call rates <97%, mismatches of phenotypic and genetic gender, heterozygosity rates +/-5 SD from the mean, and population outliers. Before imputation we reduced to SNPs genotyped on both used platforms and excluded SNPs with call rates <98%, HWE p-values <5x10-6, and minor allele frequency <1% calculated over all samples. Pre-phasing was performed with SHAPEIT v2. Imputation was performed with IMPUTE v2.3.0 using the 1000G phase1 (v3) reference panel. The statistical analysis was performed in R v2.15.2. KORA F4 including 3,080 individuals was conducted from 2006-08 as a follow-up study to KORA S4 (1999-2001). Lung function tests were performed in a random subsample of subjects born between 1946 and 1965 (age range 41–63 years). Spirometry was performed in line with the ATS/ERS recommendations using a pneumotachograph-type spirometer (Masterscreen PC, CardinalHealth, Würzburg, Germany) before and after inhalation of 200 µg salbutamol. The present study is based on maximum values of FEV1 measured before bronchodilation. The spirometer was calibrated daily using a 1L-calibration pump (CardinalHealth, Würzburg, Germany), and additionally, an internal control was used to ensure constant instrumental conditions. Genotyping was performed on the Affymetrix Axiom platform. Genotypes were called with the Affymetrix software and annotated to NCBI build 37. We excluded individuals with call rates <97%, mismatches of phenotypic and genetic gender, heterozygosity rates +/-5 SD from the mean, and population outliers. Before imputation we excluded SNPs with call rates <98%, HWE p-values <5x10-6, and minor allele frequency <1%. Pre-phasing was performed with SHAPEIT v2. Imputation was performed with IMPUTE v2.3.0 using the 1000G phase1 (v3) reference panel. The statistical analysis was performed in R v2.15.2. | The KORA authors acknowledge all members of field staffs who were involved in the planning and conduction of the KORA Augsburg studies, as well as all KORA study participants. |

| Study of Health in Pomerania (SHIP) | The Study of Health in Pomerania (SHIP) is a cross-sectional survey in West Pomerania, the north-east area of Germany [Völzke H, et al. (2011). Int J Epidemiol 40: 294-307]. A sample from the population aged 20 to 79 years was drawn from population registries. First, the three cities of the region (with 17,076 to 65,977 inhabitants) and the 12 towns (with 1,516 to 3,044 inhabitants) were selected, and then 17 out of 97 smaller towns (with less than 1,500 inhabitants), were drawn at random. Second, from each of the selected communities, subjects were drawn at random, proportional to the population size of each community and stratified by age and gender. Only individuals with German citizenship and main residency in the study area were included. Finally, 7,008 subjects were sampled, with 292 persons of each gender in each of the twelve five-year age strata. In order to minimize drop-outs by migration or death, subjects were selected in two waves. The net sample (without migrated or deceased persons) comprised 6,267 eligible subjects. Selected persons received a maximum of three written invitations. In case of non-response, letters were followed by a phone call or by home visits if contact by phone was not possible. The SHIP population finally comprised 4,308 participants (corresponding to a final response of 68.8%). Lung functions were measured during the 5-year follow-up in 1809 out of 3300 study participants. The study was approved by the ethics committee of the University of Greifswald ("Leben und Gesundheit in Vorpommern", III UV 73/01). |
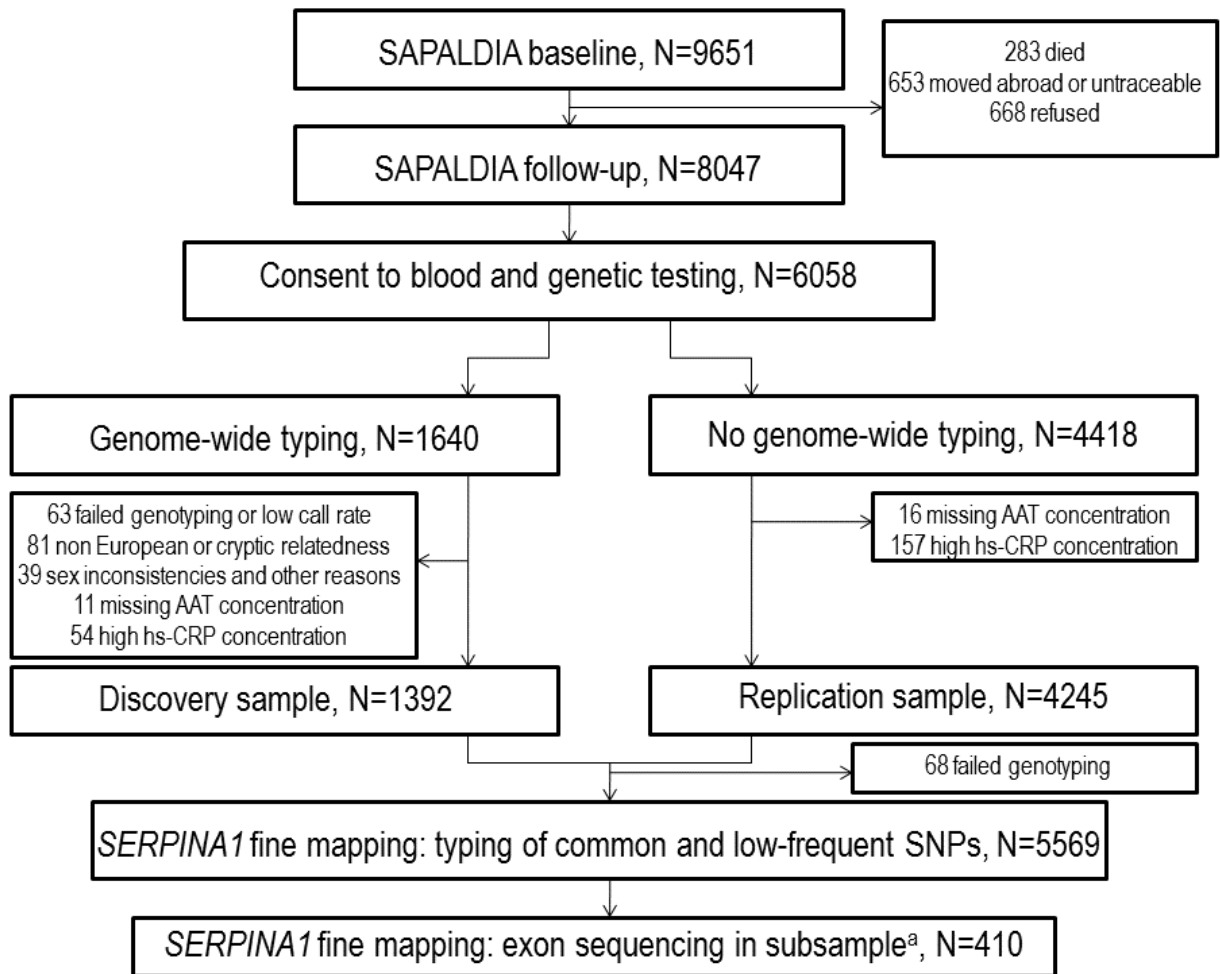
Figure S1. SAPALDIA study design for the determination of AAT associated genetic variants.
[a] consisting of subjects with abnormally low AAT levels independent of PI S or Z alleles (see Materials and Methods).

Figure S2. Q-Q plot of genome-wide -log(10) p-values for association with AAT serum level.

STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.052	-0.005	0.012		
BHS (N=1977)	0.059	0.002	0.037		
COPENHAGEN (N=6488)	0.030	-0.012	0.030		
KORA_S3 (N=536)	0.056	-0.013	0.069		
KORA_F4 (N=802)	0.055	0.070	0.067		
LBC36 (N=554)	0.063	0.102	0.082		
LHS (N=591)	0.050	0.054	0.041		
NFBC66 (N=3324)	0.018	0.004	0.017		
SAPALDIA (N=3038)	0.040	0.036	0.033		
SHIP (N=995)	0.065	0.003	0.054		
META_FIXED (N=22530)		**0.009**	**0.010**	**0.37**	**0.0**
POPULATION ISOLATES					
KORCULA (N=423)	0.060	0.094	0.088		
NSPHS (N=121)	0.036	-0.142	0.027		
ORCADES (N=288)	0.087	-0.066	0.063		
SPLIT (N=254)	0.061	-0.199	0.106		
VIS (N=537)	0.074	-0.073	0.093		
META_FIXED		**-0.114**	**0.023**	**<0.0001**	**50.7**
META_RANDOM (N=1623)		**-0.086**	**0.043**	**0.05**	

Change in FEV1 per minor allele

Figure S3. Forest plot of meta-analyzed results for the effect per minor allele of rs17580 (PI S) on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2>0.5$. Study weights (blue squares) correspond to the fixed effect meta-analyses.

STUDY_NAME	MINOR_AF	ESTIMATE	STD_ERROR	P_VALUE	I_SQUARE
GENERAL POPULATION					
B58C (N=4225)	0.020	0.050	0.038		
BHS (N=1977)	0.022	-0.075	0.066		
COPENHAGEN (N=6488)	0.027	-0.034	0.031		
KORA_S3 (N=536)	0.010	-0.229	0.148		
KORA_F4 (N=802)	0.010	-0.022	0.145		
LBC36 (N=554)	0.021	0.064	0.131		
LHS (N=591)	0.022	0.122	0.060		
NFBC66 (N=3324)	0.018	0.007	0.015		
SAPALDIA (N=3038)	0.011	0.049	0.061		
SHIP (N=995)	0.021	-0.048	0.084		
META_FIXED (N=22530)		**0.006**	**0.012**	**0.58**	**25.1**
POPULATION ISOLATES					
KORCULA (N=423)	0.017	-0.067	0.134		
NSPHS (N=121)	0.102	0.164	0.020		
ORCADES (N=288)	0.079	-0.083	0.069		
SPLIT (N=254)	0.017	0.085	0.174		
VIS (N=537)	0.026	0.071	0.134		
META_FIXED		**0.138**	**0.019**	**<0.0001**	**72.3**
META_RANDOM (N=1623)		**0.042**	**0.073**	**0.57**	

Figure S4. Forest plot of meta-analyzed results for the effect per minor allele of rs28929474 (PI Z) on FEV1 in ever-smokers, adjusted for sex, age, height and population stratification factors. Studies based on population isolates with a high degree of cryptic relatedness are presented separately. Effect estimates of meta-analyses are shown with green diamonds. I^2 is a measure of the heterogeneity between studies. Random effect meta-analyses are included if $I^2>0.5$. Study weights (blue squares) correspond to the fixed effect meta-analyses.

Perspective

Fifty Years On: GWAS Confirms the Role of a Rare Variant in Lung Disease

Alice M. Turner[1,2]*

1 University of Birmingham, QEHB research labs, Birmingham, United Kingdom, 2 Heart of England NHS Foundation Trust, Birmingham, United Kingdom

This year, 2013, is the fiftieth anniversary of the discovery of alpha one antitrypsin deficiency (AATD), a disease caused by mutation in *SERPINA1*, which predisposes to early onset lung disease, specifically chronic obstructive pulmonary disease (COPD). In this issue of *PLOS Genetics*, the timely study of Thun *et al.* [1] investigated alpha one antitrypsin (AAT) levels with a genome-wide association study (GWAS) approach in a cohort recruited primarily for studies of asthma. They demonstrate that, although a genome-wide association is detected for a common variant in *SER-PINA1*, the association disappears after adjusting the results for the presence of the known rare variants that cause AATD (PiZ and PiS). The observation of this "synthetic association" illustrates the potential for detection of rare genetic variants, with minor allele frequencies (MAF) of less than 1%, to revolutionize the understanding of pathogenesis.

Importantly, the approach used by Thun *et al.*—the sequencing of *SERPINA1*, fine mapping, and a conditional approach to statistical analysis in the regress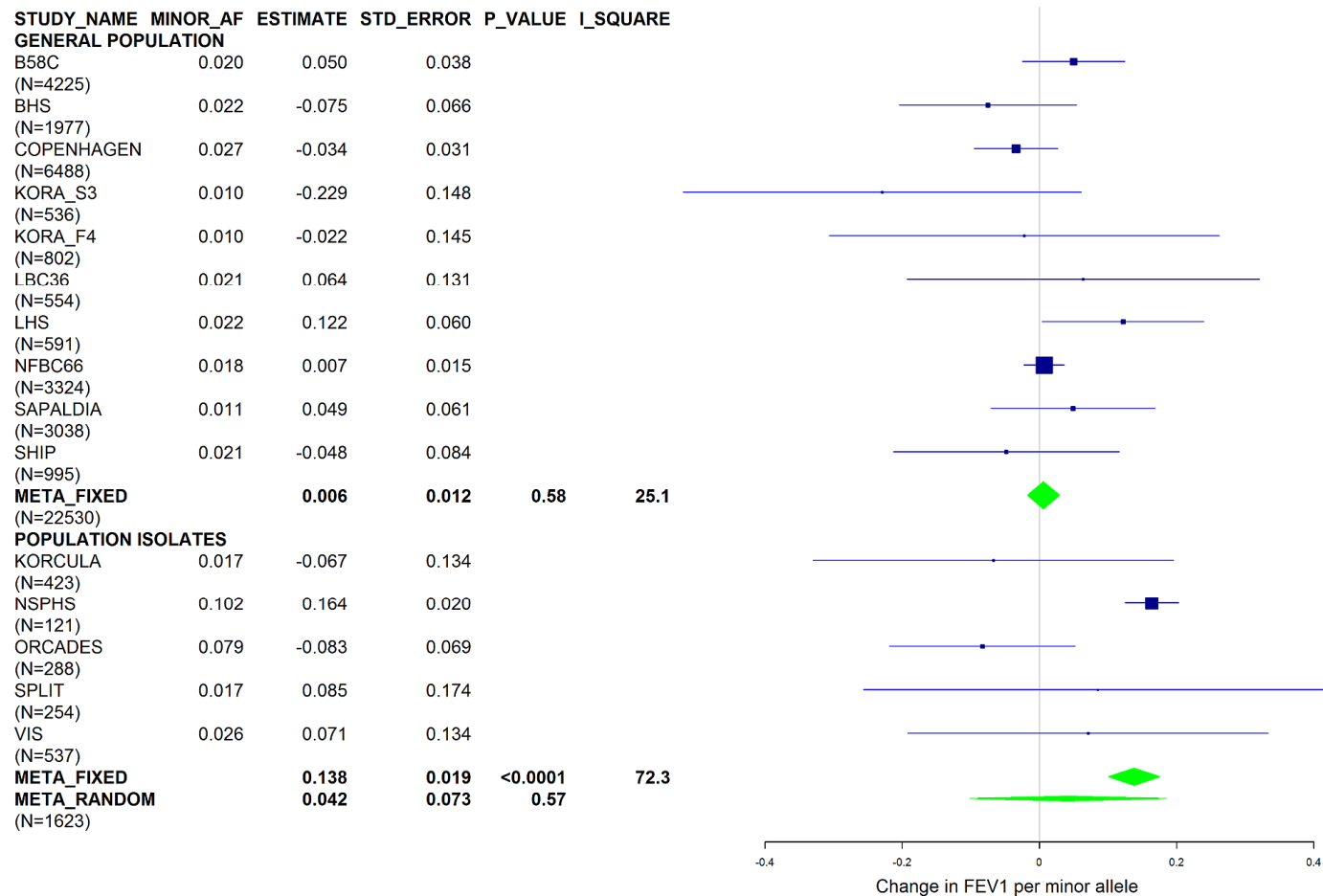ion model—does not depend on knowing the rare variants on which models need to be conditioned, and was able to reliably identify known variants with MAF 1%– 5%. It is therefore a nice demonstration of the potential for sequencing to reveal rare variants and ascertain their true contribution to traits. Thun *et al.* have also compensated for known environmental exposure that interacts with genes, such as cigarette smoke exposure ("ever smoked"), enhancing this analysis by use of a proxy measure for smoke intensity (hsCRP). The robustness of the results was demonstrated using a second large cohort. The clinical relevance of this work was demonstrated by looking at lung function as the outcome, although this revealed a more complex scenario. The results leave some unanswered questions, such as the observation that variants influencing expression which lie outside *SERPINA1* exons are not accounted for, and that recent expression quantitative trait loci (eQTL) data for *SERPINA1* did not report their associated SNPs.

Since the advent of GWAS it has become increasingly obvious that this research design cannot detect the majority of the heritability of most studied traits. Several ideas have been proposed to explain this "missing heritability," including gene–environment interaction, reduced power of GWAS to detect functional variants due to low levels of linkage disequilibrium with causative SNPs, and undetected rare variants. Thun *et al.* [1] provide an excellent realization of this last hypothesis. Most GWAS have picked up common variants that confer only a small measured increase in risk (increased odds ratios [ORs]), but have missed by design rare variants that confer larger ORs of disease. AATD and GWAS of COPD and lung function are perfect examples of this as they failed to detect both the PiZ and PiS variants, despite the well-established role of these mutations in lung disease. Only through a huge meta-analysis of over 20,000 individuals and smoking interaction modeling could the influence of PiZ [2] on lung function be detected. Thun *et al.* could use their analyses to begin to address the longstanding debate regarding the role of AATD in lung disease at the population level. Specifically, does carrying a single abnormal *SERPINA1* allele increase risk of lung disease at the population level [3]? A demonstration of an association, even at low AAT levels not considered truly deficient, would mean that AAT pathways are relevant to a far larger proportion of the population than previously thought. For this reason, the work of Thun *et al.* is not only an interesting example of synthetic association, but has potential clinical importance since it suggests that even small variation in AAT level could affect lung function [1].

Since rare variant analysis is a topic of great interest to genetic epidemiologists at present, it is worth reprising how understanding AATD has moved clinical medicine forward. It was first described in 1963 by Laurell and Eriksson, who reported an absence of the α1 band on protein electrophoresis of serum taken from a patient at a local respiratory hospital [4]; this missing band was due to very low circulating levels of AAT. The observation that people with this deficiency developed early-onset emphysema and COPD suggested a role for pathways involving AAT in pathogenesis(summarized in Figure 1). The main function of AAT is as an anti-protease, which protects tissues against neutrophil elastase (NE) [5]; the protease–antiprotease hypothesis of emphysema resulted directly from this knowledge and has driven much of the research into COPD. Clinically significant AATD-related lung disease occurs in approximately 1 in 2,500 Caucasian individuals who usually carry the PiZ or PiS variants of *SERPINA1*; importantly, both SNPs confer an increased risk of lung disease in carriers who smoke and in homozygous individuals who do not smoke [6]. However, COPD is very common, affecting up to 20% of smokers and 11% of non-smokers worldwide, and is projected to be the third most common cause of death by 2030 [7]. Consequently,

Figure 1. Simplified pathogenesis of lung disease in AATD. (1) Polymorphisms in DNA lead to structural changes in AAT which interact with (2) environmental exposure to cigarette smoke, amongst other influences. The combination of gene+environment leads to (3) neutrophilic inflammation in the lung; the main driver of disease is an inability to protect from the harmful effects of NE released by neutrophils. (4) Proteolytic destruction of lung tissue leads to the typical clinical pattern of emphysema, usually worst at the lung bases, as shown on this reconstructed image. (5 & 6) In addition, AAT polymers present in the lung, whose formation occurs due to the PiZ and to a lesser extent the PiS variant, play a smaller role in augmenting inflammation by attracting other inflammatory cells such as macrophages.
doi:10.1371/journal.pgen.1003768.g001

changing the direction of research in COPD had potential to greatly affect population health. At the time of detection of AATD, COPD as a term was not in common use; rather, patients were described as having emphysema or chronic bronchitis, both diagnosed by clinical features and chest radiography. It is now a common term and is diagnosed by a reduction in FEV1/FVC ratio on spirometry—a simple test of lung function—and may be further characterized by more complex lung function tests and computerized tomography (CT) scanning of the lungs.

Genetic research into the etiology of lung dysfunction in the general population [2] as well as the risk of COPD [8] has succeeded in identifying many variants that are significantly associated with these outcomes. Candidate gene approaches discovered other relevant proteases [9], and our increased understanding of the inflammatory cascade leading to neutrophil activation and NE release has guided the implementation of therapies targeting inflammation, the most common of which is inhaled corticosteroids (ICS). However, anti-inflammatory therapies have not been a magic bullet for COPD sufferers; indeed, the main trial seeking reduction in mortality showed only a trend in this direction (p = 0.052) when ICS were combined with long-acting beta agonists (LABA) [10]. However, lung function, flare-ups of the disease (known as exacerbations), and quality of life have been shown in many studies to improve with ICS/LABA combinations [7]. Relevant to the current paper on *SERPINA1* variants as a predictor of AAT, specific treatment for AATD has been developed in the form of AAT augmentation therapy. A meta-analysis of 1,509 patients in observational studies and randomized controlled trials (RCTs) concluded that it was beneficial, because decline in lung function was slower upon treatment [11]. More sophisticated measures of lung disease using quantitative analysis of CT scans have shown that augmentation is beneficial in the RCTs alone [12]. Whether or not the recognition that rare variants explain common variant associations at *SERPINA1* with AATD changes clinical practice remains to be seen, but the story should be a source of inspiration for genetic researchers to continue pursuing rare variants and their potential use in disease prediction and targeted therapy.

References

1. Thun GA, Imboden M, Ferrarotti I, Kumar A, Obeidat M, et al. (2013) Causal and synthetic associations of variants in the serpina gene cluster with alpha1-antitrypsin serum levels. PLoS Genet 9(8): e1003585. doi:10.1371/journal.pgen.1003585

2. Obeidat M, Wain LV, Shrine N, Kalsheker N, Artigas MS, et al. (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS One 6: e19382.

3. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, et al. (2010) alpha(1)-Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. Chest 138: 1125–1132.

4. Laurell CB, Eriksson S (1963) The electrophoretic alpha 1 globulin pattern of serum in alpha 1 antitrypsin deficiency. Scand J Clin Lab Invest 15: 132–140.

5. Carrell RW, Jeppsson JO, Laurell CB, Brennan SO, Owen MC, et al. (1982) Structure and variation of human alpha 1-antitrypsin. Nature 298: 329–334.

6. Luisetti M, Seersholm N (2004) Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax 59: 164–169.

7. Global Initiative for Obstructive Lung Disease.A-Available: http://www.goldcopd.org. Accessed June 2013.

8. 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.

9. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, et al. (2009) MMP12, lung function, and COPD in high-risk populations. N Engl J Med 361: 2599–2608.

10. Calverley PM, Anderson JA, Celli B, Ferguson GT, Jenkins C, et al. (2007) Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med 356: 775–789.

11. Chapman KR, Stockley RA, Dawkins C, Wilkes MM, Navickis RJ (2009) Augmentation therapy for alpha1 antitrypsin deficiency: a meta-analysis. COPD 6: 177–184.

12. Stockley RA, Parr DG, Piitulainen E, Stolk J, Stoel BC, et al. (2010) Therapeutic efficacy of alpha-1 antitrypsin augmentation therapy on the loss of lung tissue: an integrated analysis of 2 randomised clinical trials using computed tomography densitometry. Respir Res 11: 136.

5.3   Paper 4: *SERPINA1* PiZ and PiS Heterozygotes and Lung Function Decline in the SAPALDIA Cohort.

This paper was published:

PLOS ONE

SERPINA1 PiZ and PiS Heterozygotes and Lung Function Decline in the SAPALDIA Cohort

Gian-Andri Thun[1,2], Ilaria Ferrarotti[3], Medea Imboden[1,2], Thierry Rochat[4], Margaret Gerbase[4], Florian Kronenberg[5], Pierre-Olivier Bridevaux[4], Elisabeth Zemp[1,2], Michele Zorzetto[3], Stefania Ottaviani[3], Erich W. Russi[6], Maurizio Luisetti[3], Nicole M. Probst-Hensch[1,2]*

1 Swiss Tropical and Public Health Institute, Basel, Switzerland, 2 University of Basel, Basel, Switzerland, 3 Center for Diagnosis of Inherited Alpha1-antitrypsin Deficiency, Institute for Respiratory Disease, IRCCS San Matteo Hospital Foundation, University of Pavia, Pavia, Italy, 4 Division of Pulmonary Medicine, University Hospital of Geneva, Geneva, Switzerland, 5 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria, 6 Pulmonary Division, University Hospital of Zurich, Zurich, Switzerland

Abstract

Background: Severe alpha1-antitrypsin (AAT) deficiency is a strong risk factor for COPD. But the impact of gene variants resulting in mild or intermediate AAT deficiency on the longitudinal course of respiratory health remains controversial. There is indication from experimental studies that pro-inflammatory agents like cigarette smoke can interact with these variants and thus increase the risk of adverse respiratory health effects. Therefore, we tested the effect of the presence of a protease inhibitor (Pi) S or Z allele (PiMS and PiMZ) on the change in lung function in different inflammation-exposed subgroups of a large, population-based cohort study.

Methodology and Principal Findings: The SAPALDIA population includes over 4600 subjects from whom *SERPINA1* genotypes for S and Z alleles, spirometry and respiratory symptoms at baseline and after 11 years follow-up, as well as proxies for inflammatory conditions, such as detailed smoking history, obesity and high sensitivity C-reactive protein (hs-CRP), were available. All analyses were performed by applying multivariate regression models. There was no overall unfavourable effect of PiMS or PiMZ genotype on lung function change. We found indication that PiZ heterozygosity interacted with inflammatory stimuli leading to an accelerated decline in measures in use as indices for assessing mild airway obstruction. Obese individuals with genotype PiMM had an average annual decline in the forced mid expiratory flow (ΔFEF25-75%) of 58.4 ml whereas in obese individuals with PiMZ it amounted to 92.2 ml (p = 0.03). Corresponding numbers for persistent smokers differed even more strongly (66.8 ml (PiMM) vs. 108.2 ml (PiMZ), p = 0.005). Equivalent, but less strong associations were observed for the change in the FEV1/FVC ratio.

Conclusions: We suggest that, in addition to the well established impact of the rare PiZZ genotype, one Z allele may be sufficient to accelerate lung function decline in population subgroups characterized by elevated levels of low grade inflammation.

Introduction

Reduced lung function measures are the most common diagnostic parameter to detect airway obstruction which is the main characteristic of asthma and chronic obstructive pulmonary disease (COPD). The only functionally characterized gene variants associated with airway obstruction and increased loss of lung

function are infrequent polymorphisms in the *SERPINA1* gene causing deficiency of alpha1-antitrypsin (AAT) [1]. This antiprotease inhibits neutrophil elastase, an enzyme that degrades pulmonary elastic fibers. Homozygosity for the protease inhibitor deficiency variant Z (PiZZ, also referred to as severe AAT deficiency) and compound heterozygosity for both deficiency variants S and Z (PiSZ) are widely accepted risk factors for airway obstruction and accelerated lung function decline, particularly among smokers [2]. But since these allele combinations all have frequencies below 0.1% in the general European population [3], they only account for 2-5% of all COPD cases. The more prevalent heterozygous genotypes PiMS and PiMZ (M stands for the wildtype allele) reduce the AAT blood levels only slightly [4] and are therefore referred to as mild (for PiMS) and intermediate (for PiMZ) AAT deficiencies.

While PiMS is generally believed not to be associated with low lung function or a higher risk of COPD [5], the evidence for PiMZ remains unclear even in the light of a meta-analysis [6]. The few population-based longitudinal studies have not shown adverse health effects, but they varied with regard to the phenotype studied and the inclusion of gene-environment interactions [7,8,9]. An investigation restricted to smokers showed that PiMZ was overrepresented in the group with rapid FEV1 (forced expiratory volume in one second) decline, suggesting that susceptibility may be refined to population subgroups with elevated inflammatory and proteolytic stress in the lungs [10]. These processes may locally increase cleavage, as well as oxidant-induced inactivation [11] and polymerization [12] of AAT, leading to a further reduction of this enzyme in PiMZ carriers. Apart from inhalant triggers like smoking, systemic inflammation may also compromise pulmonary health [13,14] and therefore particularly affect individuals with reduced AAT levels.

We hypothesized therefore that the intrapulmonary antiproteolytic capacity in people with mild or intermediate AAT deficiency may not be sufficient to counterbalance an excess of inflammatory triggers targeting the respiratory system. We used the SAPALDIA cohort (Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults) to test such possibly unfavourable effects of PiS or PiZ heterozygosity on the longitudinal course of lung function over 11 years of follow-up in the general population. The large and well characterized study population allowed us to particularly study subgroups exposed to elevated local airway or systemic inflammatory conditions, such as active and passive smokers, and people suffering from obesity.

Results

Study Sample

A comparison of characteristics between SAPALDIA follow-up participants included and excluded in this study is provided in Table 1. The study population consisted of a healthier sample with a higher percentage of never-smokers and fewer obese individuals. A comparison between the three genotype classes showed highly significant differences in AAT serum levels (Table 1). Compared to PiM homozygotes, unadjusted AAT serum concentrations were more than 16% and 38% reduced in PiS and PiZ heterozygotes, respectively. There were no differences in circulating high-sensitivity C-reactive protein (hs-CRP), the main marker for systemic inflammation, between the genotype groups. Furthermore, the genotype classes differed slightly in the unadjusted forced expiratory flow over the middle half of FVC (FEF25-75%) at baseline and in unadjusted declines of FEF25-75%.

Adjusted Spirometric Decline Rates according to Genotype

In adjusted models, neither PiMS nor PiMZ subjects exhibited statistically significant steeper annual declines than PiMM individuals in any measure of lung function (all p≥0.07, Table 2 and Table 3). Hypothesizing that PiMS and PiMZ carriers can only compensate their reduced anti-proteolytic capacity in pulmonary tissue if no excess of pulmonary or systemic inflammation is present, we tested if stratification by smoking or obesity status may alter these associations. While we could not find any significant association between the presence of a S or Z allele and ΔFEV1 or ΔFVC (forced vital capacity) irrespective of the smoking or obesity category (Table 2), smokers and obese individuals with PiMZ genotype showed elevated declines in FEF25-75% (Table 3). In ever smokers, PiMZ carriers lost on average additional 17.4 ml per year compared to PiMM in FEF25-75% (p = 0.05), and this difference became more pronounced in people who smoked at baseline and follow-up (41.4 ml in persistent smokers, p = 0.005). A similar pattern could be observed in obese participants (ΔFEF25-75% = 58.4 ml (PiMM) vs. 92.2 ml (PiMZ), p = 0.03). There was no such effect in never-smokers exposed to environmental tobacco smoke. Values for Δ (FEV1/FVC) consistently confirmed this trend, but associations did not reach statistical significance. For ΔFEF25-75%, the presence of a Z allele interacted statistically significant with smoking (p$_{interaction}$ = 0.04 with smoking status (persistent vs. never) and p$_{interaction}$ = 0.002 with packyears between the two surveys), but not with obesity status (p = 0.14). Statistically significant modification of the Z allele effect on Δ (FEV1/FVC) could be observed for packyears between the two surveys and obesity status (p$_{interaction}$ = 0.04 and 0.08, respectively). As we had previously found sex differences in the association between circulating AAT concentrations and lung function [15], we further evaluated a possible effect modification by sex. However, sex did not significantly modify the allele effect on any lung function measure (all p$_{interaction}$ ≥ 0.15).

High-sensitivity C-reactive Protein (hs-CRP) as a Proxy for Low Systemic Inflammation

In order to strengthen the hypothesis that low grade systemic inflammation drove the observed Z allele effect on pulmonary function, we investigated the associations in participants with elevated hs-CRP levels at follow-up (defined as the upper tertile of the study population, i.e.≥1.8 mg/l). Consistent with our findings for smokers and obese people, declines in PiZ heterozygotes were significantly enhanced for FEF25-75% (99.3 ml (PiMZ) vs. 71.2 ml (PiMM), p = 0.006) and suggestively for FEV1/FVC (5.46‰ (PiMZ) vs. 4.00‰ (PiMM), p = 0.07) (Table 3), but not for FEV1 or FVC. Interestingly, also PiS heterozygotes in the upper tertile of hs-CRP values showed statistically significant larger declines in FEF25-75% (p = 0.003) and FEV1/FVC (p = 0.04) than PiMM subjects.

Incidence of Airway Obstruction and Respiratory Symptoms

As a next step we tried to expand our analyses on the change of spirometry measures with clinically relevant outcomes like the incidence of airway obstruction or the development of respiratory symptoms. 103 PiMZ carriers did not show airway obstruction at baseline which was defined by FEV1/FVC < 0.7. In 17 of them a smaller ratio than 0.7 was calculated at follow-up, which classified them as incident cases. Comparing with PiMM individuals this resulted in an adjusted odds ratio of 1.11 (95% CI = 0.64 to 1.93, p = 0.70; Table 4). Corresponding investigations with individuals exposed to elevated inflammatory conditions did not show an

Table 1. Characteristics of SAPALDIA follow-up participants at baseline (1991; upper part), follow-up (2002; middle part), and between the two examinations (lower part).

	Not included in Study (N = 3372)	Included in Study (N = 4675)	PiMM (N = 4207)	PiMS (N = 356)	PiMZ (N = 112)	P-value[a]
At baseline						
% women	51.2	52.4	52.6	50.8	53.6	0.80
Age (y), median, IQR	42.4 (32.6, 51.1)	41.4 (31.7, 49.7)	41.5 (31.6, 49.8)	41.3 (32.9, 49.6)	38.3 (31.3, 47.3)	0.33
% never smokers	39.9 (N = 3360)	49.6	49.3	51.7	54.5	0.39
% current smokers	34.9 (N = 3360)	29.6	30.0	27.2	24.1	0.24
% ETS in never smokers	30.6 (N = 1340)	28.3 (N = 2318)	28.5 (N = 2073)	24.5 (N = 184)	34.4 (N = 61)	0.29
BMI (kg/m^2), median, IQR	23.7 (21.3, 26.3; N = 3306)	23.1 (21.1, 25.6)	23.1 (21.0, 25.7)	23.3 (21.2, 25.7)	22.9 (21.3, 24.9)	0.80
% obese[b]	7.6 (N = 3306)	5.3	5.4	4.5	4.5	0.72
FEV1 (mL), mean±SD	3485±864 (N = 2910)	3576±824	3570±825	3617±819	3627 ± 787	0.35
FVC (mL), mean±SD	4415±1062 (N = 2929)	4527±1022	4522±1025	4580±1000	4553±987	0.57
FEF25-75% (mL), mean±SD	3371±1228 (N = 2794)	3436±1223	3426±1221	3468±1212	3706±1320	0.05
FEV1/FVC (%), mean±SD	79.10±7.93 (N = 2794)	79.21±7.59	79.18±7.56	79.07±7.83	80.61±8.01	0.14
At follow-up						
% never smokers	40.7 (N = 2936)	48.1	47.8	49.7	53.6	0.40
% current smokers	32.3 (N = 2936)	22.6	22.8	21.9	16.1	0.23
% ETS in never smokers	15.1 (N = 1189)	15.4 (N = 2249)	15.5 (N = 2012)	14.7 (N = 177)	15.0 (N = 60)	0.96
BMI (kg/m^2), median, IQR	25.7 (23.0, 28.8; N = 1923)	25.2 (22.7, 28.2)	25.2 (22.7, 28.2)	25.4 (22.8, 28.4)	25.2 (23.1, 27.6)	0.88
% obese[b]	18.1 (N = 1923)	15.2	15.2	15.2	14.3	0.96
AAT (g/L), mean±SD	1.25±0.22 (N = 1642)	1.27±0.20	1.30±0.19	1.09±0.16	0.80±0.11	<0.001
hs-CRP (mg/L), median, IQR	1.1 (0.5, 2.3; N = 1642)	1.0 (0.5, 2.3)	1.0 (0.5, 2.3)	1.0 (0.5, 2.0)	0.9 (0.5, 2.5)	0.16
Between baseline and follow-up						
ΔFEV1 (mL/y), mean±SD	-35.2±32.6 (N = 1278)	-35.3±29.8	-35.1±29.8	-37.7±30.1	-36.3±28.7	0.27
ΔFVC (mL/y), mean±SD	-26.1±43.7 (N = 1266)	-24.1±40.2	-24.1±40.2	-25.0±39.7	-22.4±41.5	0.82
ΔFEF25-75% (mL/y), mean±SD	-64.2±64.1 (N = 1070)	-71.1±65.2	-70.4±64.8	-76.0±64.9	-84.3±77.9	0.03
Δ (FEV1/FVC) (‰/y), mean±SD	-3.49±5.19 (N = 1071)	-4.05±4.95	-4.01±4.94	-4.26±5.01	-4.62±4.99	0.32

AAT: Alpha1-antitrypsin; BMI: Body mass index; ETS: regularly exposed to environmental tobacco smoke; FEF25-75%: mid expiratory flow; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; hs-CRP: high-sensitivty C-reactive protein; IQR: Interquartile range; SD: Standard deviation.
[a]Pearson's $\chi 2$ test for testing equal proportions, analysis of variance (ANOVA) for testing equal means of normally distributed data (AAT levels and spirometric measurements), and Kruskal-Wallis test for testing equal distributions in not normally distributed data (age and BMI values) between the three genotype classes.
[b]Obese subjects were defined as having a BMI≥30kg/m^2.
doi:10.1371/journal.pone.0042728.t001

increased obstruction risk for Z heterozygotes, but were based on very few numbers of cases. As far as the development of respiratory symptoms like regular cough, phlegm or shortness of breath during sleep was concerned, there seemed to be a trend that PiMZ individuals were more susceptible than PiMM carriers, in particular in subgroups exposed to pro-inflammatory conditions (Table 4). But the limited statistical power did unfortunately not allow a comprehensive evaluation of our hypothesis.

Validity

As chronic asthma is suggested to be associated with accelerated loss of pulmonary function [16] and in the absence of post-bronchodilation spirometry, we conducted a sensitivity analysis by including only subjects who did not report physician-diagnosed asthma. Furthermore, level of lung function change in adults is also determined by lung function growth during early adulthood, and to account for that, we performed another sensitivity analysis by only including participants older than 30 years of age at baseline. Neither of the two restrictions did essentially alter the results (Table S1).

Replacing the change of FEF25-75% by the change of the ratio FEF25-75%/FVC led to very similar conclusions (Table S2). Finally, in order to detect potential participation bias, we weighted each observation inverse to the probability of being included in the study sample. None of the results of the regression analyses did materially change (Table S3).

Discussion

In the present study, neither PiS nor PiZ heterozygosity influenced longitudinal lung function measured by ΔFEV1 or ΔFVC, independent of smoking or obesity status. However, PiMZ genotype was associated with an accelerated FEF25-75% decline in smoking and obesity subgroups from the general population. Results for participants in the upper tertile of hs-CRP values strengthened the notion that PiMZ carriers might be more susceptible to systemic pro-inflammatory conditions with respect to lung function parameters indicating narrowing of small airways.

Table 2. Adjusted mean values in ΔFEV1 and ΔFVC over 11 years of follow-up comparing different *SERPINA1* genotypes.

	ΔFEV1(ml/y)	95%CI	P-value	ΔFVC(ml/y)	95%CI	P-value
Total						
PiMM, N = 4207	-35.2	-36.0 to -34.4		-24.1	-25.2 to -23.0	
PiMS, N = 356	-36.9	-39.7 to -34.1	0.24	-24.7	-28.3 to -21.0	0.77
PiMZ, N = 112	-36.0	-41.0 to -31.1	0.74	-23.3	-29.8 to -16.7	0.80
Passive Smokers[a]						
PiMM, N = 691	-32.3	-34.5 to -30.2		-21.4	-24.1 to -18.6	
PiMS, N = 52	-33.9	-41.5 to -26.3	0.69	-26.3	-35.9 to -16.7	0.33
PiMZ, N = 22	-41.3	-52.9 to -29.7	0.14	-34.5	-49.2 to -19.7	0.09
Ever Smokers						
PiMM, N = 2194	-35.1	-36.4 to -33.8		-25.0	-26.7 to -23.2	
PiMS, N = 179	-36.8	-40.8 to -32.7	0.43	-24.1	-29.5 to -18.6	0.75
PiMZ, N = 52	-35.4	-42.8 to -27.9	0.94	-21.6	-31.6 to -11.6	0.51
Persistent Smokers[b]						
PiMM, N = 922	-35.6	-39.3 to -31.9		-26.6	-31.3 to -21.9	
PiMS, N = 74	-41.7	-48.7 to -34.8	0.07	-30.6	-39.4 to -21.7	0.35
PiMZ, N = 18	-41.2	-54.5 to -27.9	0.40	-24.0	-40.8 to -7.1	0.75
Obese subjects[c]						
PiMM, N = 653	-36.8	-42.4 to -31.3		-33.6	-41.1 to -26.1	
PiMS, N = 55	-37.1	-46.3 to -28.0	0.94	-29.7	-42.1 to -17.4	0.48
PiMZ, N = 16	-48.0	-62.9 to -33.1	0.12	-34.7	-54.8 to -14.6	0.91

CI: confidence interval. FEV1: forced expiratory volume in 1 second. FVC: forced vital capacity.
Covariates included sex, linear and squared age, recruiting area, smoking history (packyears at baseline, as well as linear and squared packyears between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up.
[a]Passive smokers were defined as never smokers who declared regular exposure to environmental tobacco smoke within one year prior to the baseline or follow-up examination.
[b]Persistent smokers were classified as subjects who declared current smoking at both examinations.
[c]Obese subjects were defined as BMI≥30kg/m² at the baseline or follow-up examination.
doi:10.1371/journal.pone.0042728.t002

The Role of Smoking and Pulmonary Oxidative Stress

It is well established that smokers suffering from severe AAT deficiency, a condition accompanied by only 15% of normal AAT blood concentrations, are particularly vulnerable to developing early onset COPD [2]. There is less evidence that intermediate or even mild AAT deficiency modify the effect of smoking or other oxidative inhalants on lung function. In a small study of 56-year-old men, a higher mean annual decrease in FEV1 in smoking PiMZ individuals was reported as compared with non-smoking PiMZ or smoking PiMM individuals [17], but larger studies could not find such an interaction [7,9]. Results of studies investigating the impact of environmental and occupational exposure are heterogeneous [18]. Passive smoking has been shown in school children to be associated with cross-sectional lower lung function only in children having low levels of AAT in the blood, and particularly in measures of mid- to end-expiratory flow rates [19]. A recent longitudinal study found accelerated lung function declines in New York City firefighters of PiMZ genotype compared to PiMM in the years after World Trade Center collapse accompanied by massive air pollution, and no such difference could be observed prior to September 11, 2001 [20].

Support for a gene-environment interaction comes also from genome-wide as well as from experimental studies. While genome-wide association studies on COPD and lung function have not found the *SERPINA1* locus among the top hit signals [21,22,23], it was the most strongly associated candidate gene in ever smokers in a comprehensive evaluation of potential lung function associated

genes in more than 20,000 individuals from the general population [24]. Oxidation of the Z form by cigarette smoke induced its polymerization in lung tissue of transgenic PiZZ mice [12]. Such polymers could also be detected on the lung surface in human PiZ homozygotes, originating either from lung epithelial cells or macrophages and independent of the main hepatocytic secretion [25]. In addition to being an ineffective elastase inhibitor, these polymers attracted a higher number of neutrophils, further shifting the protease-antiprotease equilibrium towards proteolysis. As subjects with PiMZ genotype also exhibit elevated levels of intrapulmonary neutrophilic inflammation [26], they are potentially more susceptible to oxidative inhalants. The observed dose-response trend in the interaction between smoking and the presence of a Z allele is in agreement with this biological concept.

The Role of Obesity and Systemic Low Grade Inflammation

The observed accelerated loss of pulmonary function in obese subjects carrying the PiMZ genotype is novel. The often observed inverse association between obesity and lung function is in part explained by a mechanical effect of obesity on lung volume and airway caliber. Yet more limited evidence suggests an additional effect of obesity on peripheral airway obstruction that may be related to systemic low grade inflammation [27]. Adipose tissue from subjects who are overweight or obese produces pro-inflammatory adipokines that spill over to the blood stream. Elevated circulating concentrations of inflammation markers such

Table 3. Adjusted mean values in Δ (FEV1/FVC) and ΔFEF25-75% over 11 years of follow-up comparing different *SERPINA1* genotypes.

	ΔFEF25-75%(ml/y)	95%CI	P-value	Δ (FEV1/FVC)(‰/y)	95%CI	P-value
Total						
PiMM, N = 4207	-70.6	-72.5 to -68.7		-4.03	-4.17 to -3.89	
PiMS, N = 356	-74.4	-80.8 to -68.0	0.26	-4.14	-4.62 to -3.66	0.65
PiMZ, N = 112	-81.4	-92.8 to -70.0	0.07	-4.46	-5.31 to -3.60	0.33
Passive Smokers[a]						
PiMM, N = 691	-67.5	-72.4 to -62.5		-3.88	-4.23 to -3.53	
PiMS, N = 52	-65.5	-82.8 to -48.3	0.84	-3.01	-4.23 to -1.79	0.18
PiMZ, N = 22	-65.8	-92.3 to -39.4	0.91	-3.11	-4.98 to -1.24	0.43
Ever Smokers						
PiMM, N = 2194	-70.0	-73.0 to -67.1		-3.84	-4.06 to -3.62	
PiMS, N = 179	-74.9	-84.2 to -65.6	0.32	-4.16	-4.85 to -3.47	0.38
PiMZ, N = 52	-87.4	-104.5 to -70.4	0.05	-4.80	-6.07 to -3.53	0.14
Persistent Smokers[b]						
PiMM, N = 922	-66.8	-75.2 to -58.4		-3.90	-4.53 to -3.27	
PiMS, N = 74	-74.0	-89.8 to -58.3	0.34	-4.62	-5.81 to -3.44	0.21
PiMZ, N = 18	-108.2	-138.1 to -78.2	0.005	-5.25	-7.51 to -2.99	0.23
Obese Subjects[c]						
PiMM, N = 653	-58.4	-70.1 to -46.7		-2.81	-3.70 to -1.93	
PiMS, N = 55	-59.5	-78.8 to -40.1	0.90	-3.61	-5.07 to -2.15	0.22
PiMZ, N = 16	-92.2	-123.7 to -60.6	0.03	-5.13	-7.52 to -2.75	0.05
Subjects in upper tertile of hs-CRP[d]						
PiMM, N = 1387	-71.2	-74.8 to -67.6		-4.00	-4.28 to -3.71	
PiMS, N = 99	-89.8	-102.0 to -77.7	0.003	-5.03	-5.99 to -4.06	0.04
PiMZ, N = 36	-99.3	-119.3 to -79.4	0.006	-5.46	-7.04 to -3.88	0.07

CI: confidence interval. FEF25-75%: forced mid expiratory flow. FEV1: forced expiratory volume in 1 second. FVC: forced vital capacity.
Covariates included sex, linear and squared age, recruiting area, smoking history (packyears at baseline, as well as linear and squared packyears between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up.
[a]Passive smokers were defined as never smokers who declared regular exposure to environmental tobacco smoke within one year prior to the baseline or follow-up examination.
[b]Persistent smokers were classified as subjects who declared current smoking at both examinations.
[c]Obese subjects were defined as BMI≥30kg/m² at the baseline or follow-up examination.
[d]Corresponded to a level of ≥ 1.8 mg/l.
doi:10.1371/journal.pone.0042728.t003

as hs-CRP or interleukin-6 (IL-6) are both, higher in obese persons [28,29] and associated with accelerated lung function decline [30,31]. Also in the SAPALDIA cohort, we have previously reported that accelerated lung function decline and obesity are both associated with increased hs-CRP, particularly in women [32]. Mendelian randomization approaches to examine causality between increased hs-CRP and lower respiratory function led to conflicting results [30,33]. However, these studies did not assess the effect of *CRP* gene variants on lung function separately in groups exposed to pro-inflammatory agents such as smokers.

This Study in Context

A longitudinal study in smokers reported an overrepresentation of PiMZ participants in rapid lung function decliners [10]. However, population-based cohort studies did not find consistent PiMZ effects on lung function decline, neither generally, nor in smoking strata [7,8,9]. The only marginally statistical significant result was found in non-smoking PiMZ carriers who showed a steeper unadjusted lung function decline than non-smoking PiMM subjects [7]. Our study was generally in good agreement with these

results as we could not detect any statistical significant effect of *SERPINA1* genotypes on **Δ**FEV1, which was the primary focus of the previously mentioned studies. Compared to those existing publications we included in our analysis a wider range of spirometry measures and found associations with the decline of FEF25-75% as well as weaker, but relatively consistent associations with the decline of the FEV1/FVC ratio. Both these two measures are in use to assess early airway obstruction and may probably best reflect the volume of the small airways [34,35]. There are hardly any large studies which used these measures in connection with *SERPINA1* genotypes so far, apart from a recent study of two large populations that found PiMZ genotypes associated with lower FEV1/FVC ratio and with more severe emphysema on chest computer tomography scan, but not with COPD status [36]. Flow related spirometric characteristics such as FEF25-75% may be decreased in the presence of airway abnormalities including inflammation or alterations in elastic recoil, two important correlates of AAT deficiency [37]. For example, interactions between glutathione S-transferase (GST) deficiency genotypes and passive smoking were strongest for mid expiratory flow measures

Table 4. Adjusted odds ratios for developing airway obstruction (FEV1/FVC<0.7) and respiratory symptoms over 11 years of follow-up comparing different *SERPINA1* genotypes.

	Incidence of Airway Obstruction				Incidence of Respiratory Symptoms			
	N (cases/total)	OR	95%CI	P-value	N (cases/total)	OR	95%CI	P-value
Total								
PiMM	601/3775	ref.			380/3547	ref.		
PiMS	50/315	0.97	0.70 to 1.35	0.86	33/307	1.07	0.73 to 1.57	0.73
PiMZ	17/103	1.11	0.64 to 1.93	0.70	15/101	1.52	0.86 to 2.69	0.15
Ever Smokers								
PiMM	346/1920	ref.			240/1780	ref.		
PiMS	30/152	1.12	0.72 to 1.74	0.62	18/144	0.99	0.58 to 1.66	0.96
PiMZ	9/47	1.20	0.56 to 2.60	0.64	10/43	1.94	0.93 to 4.06	0.08
Persistent Smokers[a]								
PiMM	162/801	ref.			131/705	ref.		
PiMS	14/63	0.90	0.46 to 1.78	0.76	8/56	0.76	0.34 to 1.68	0.50
PiMZ	1/17	0.29	0.04 to 2.33	0.25	3/13	1.19	0.31 to 4.56	0.80
Obese Subjects[b]								
PiMM	94/555	ref.			60/514	ref.		
PiMS	6/47	0.63	0.25 to 1.62	0.34	6/43	1.26	0.49 to 3.22	0.63
PiMZ	4/15	1.98	0.57 to 6.83	0.28	4/16	3.24	0.95 to 11.07	0.06
Subjects in upper tertile of hs-CRP[c]								
PiMM	229/1200	ref.			140/1122	ref.		
PiMS	18/88	1.01	0.57 to 1.79	0.96	13/84	1.35	0.71 to 2.56	0.35
PiMZ	7/34	1.04	0.42 to 2.55	0.93	8/32	2.51	1.07 to 5.85	0.03

CI: confidence interval. hs-CRP: high-sensitivity C-reactive protein. OR: odds ratio.
Covariates included sex, linear and squared age, recruiting area, smoking history (packyears at baseline, as well as linear and squared packyears between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up.
[a]Persistent smokers were classified as subjects who declared current smoking at both examinations.
[b]Obese subjects were defined as BMI≥30kg/m^2 at the baseline or follow-up examination.
[c]Corresponded to a level of ≥ 1.8 mg/l.
doi:10.1371/journal.pone.0042728.t004

in children [38]. However, these measures have often been criticized for being more variable and therefore less reliable than FEV1 [39]. We observed a correlation coefficient of 0.82 between baseline and follow-up FEF25-75% in SAPALDIA which is smaller than the one for FEV1 (0.92) and FVC (0.91), but larger than the one for the more commonly used FEV1/FVC ratio (0.74). Since we consistently found main and interacting effects of air pollution strongest for this mid flow parameter [40,41,42], it seems unlikely that the results of the present study are driven by measurement error. Moreover, FEF25-75% was found to have a high heritability in families with severe COPD [37].

Strengths and Weaknesses

The strength of this study is its large sample size, its detailed characterization of subjects, and its stringent quality control of spirometry [43]. The credibility of our results is supported by the fact that the reduction of AAT serum levels in PiMS and PiMZ compared to PiMM subjects was similar to that described by others [7]. Compared to the hitherto existing publications, we carefully excluded carriers of additional, rare mutations influencing AAT serum levels in order to diminish misclassification of wildtype alleles.

Our study has some limitations. First, a possible selection of healthy individuals may limit the generalizability of the results. However, giving more weight to underrepresented groups within

the study sample did not alter the results. Moreover, if persons with low levels of lung function were preferentially lost among PiMZ carriers, stated effects may be an underestimation of the true effect. Second, PiMZ individuals showed slightly higher baseline FEF25-75% and FEV1/FVC values which can be partially explained by the younger age and the reduced number of smokers in this group, but which may question the clinical relevance of the accelerated decline in these measures. Yet, the combination of a higher level of cross-sectional lung function and a steeper lung function decline after exposure to inflammatory agents parallels observations in New York City firefighters before and after the September 11 attacks [20]. Furthermore, spirometric measurements were carried out without bronchodilator [44], which hinders a clinically acceptable definition of airway obstruction. Unlike some comparable studies [9,10], we performed spirometry at only two time points. This makes our change values susceptible to imprecision, but we do not expect substantial measurement error for several reasons. The same spirometers and stringent quality criteria at baseline and follow-up were applied, correlation coefficients between the measurements at the two time points were high, and the direction of most genotype effects were consistent for all lung function outcomes. In addition, since it would be unlikely that any measurement error is associated with the *SERPINA1* genotype, misclassification would be non-differential, which indicates an underestimation rather than an overestimation of

the real effect. Another limitation is the low statistical power for analyzing obese or persistent smoking PiMZ carriers despite the large cohort size. Therefore, our findings must be interpreted with caution. If a correction for multiple testing was applied, none of the results would remain statistically significant (since all observed p-values > 0.001). For the same reason, we could neither distinguish emerging from persistent obesity, nor could we form strata with enhanced hs-CRP levels of clinical relevance (i.e.>10 mg/l). Finally, the clinical meaning of the observed excess decline in PiZ heterozygotes could not be reliably estimated, as the numbers of incident cases with airway obstruction or respiratory symptoms were too small to investigate differences between *SERPINA1* genotype classes with respect to inflammatory conditions. Nevertheless, we found indication that those subgroups are more likely to develop respiratory symptoms and we know from literature that mid expiratory flow rates have been shown to be a powerful predictor of mortality from COPD, independent of FEV1 [45].

Conclusion

We confirm in this population-based study that neither PiMS, nor PiMZ carriers have a substantial impact on longitudinal lung function. There is indication however, that the presence of one Z allele may be sufficient to accelerate loss of small airway volume and incidence of respiratory symptoms in defined population subgroups which are exposed to pro-inflammatory agents and conditions. This is a potentially relevant observation as the prevalence of PiMZ genotype is quite common in Western Europe [3] and was 2.4% in this study sample. In order to estimate the public health relevance of our findings, future studies must associate subgroups of PiMZ individuals to post-bronchodilator lung function and clinically relevant outcomes like respiratory symptoms, emphysema by chest computer tomography and hospitalization for COPD.

Materials and Methods

Ethics Statement

SAPALDIA was approved by the Swiss Academy of Medical Sciences, the supraregional ethics committee for clinical research (UREK, Project Approval Number 123/00) and the Cantonal Ethics Committees for each of the eight examination areas (Ethics commissions of the cantons Aargau, Basel, Geneva, Grisons, Ticino, Valais, Vaud and Zurich). Participants were required to give written consent before any part of the health examination was conducted either globally (for all health examinations) or separately for each investigation.

Study Population

In 1991, a random sample of 9651 adults, aged 18-60 years, from eight areas in Switzerland underwent a detailed health examination including a questionnaire about respiratory health, occupational and lifestyle exposures [46]. Participants were predominantly of European-Caucasian ethnicity and represented urban and rural areas. Eleven years later, 8047 persons were reassessed [47]. 6058 follow-up subjects provided blood samples and consented to DNA analysis. 5274 of these subjects underwent spirometry testing at baseline and follow-up. Not included in this analysis were participants with missing smoking history or body mass index (BMI) data (n = 525), subjects without valid hs-CRP (n = 18), and subjects for whom genotyping of the S or Z allele either failed (n = 6) or resulted in PiS homozygosity (n = 10), PiSZ compound heterozygosity (n = 10) or PiZ homozygosity (n = 1).

Other *SERPINA1* rare mutations which lower AAT blood levels were detected according to a procedure described elsewhere [48] in additional 29 samples which were also excluded. Our study sample included thus 4675 subjects.

Measurements

Spirometry was assessed according to American Thoracic Society criteria using the same spirometers in 1991 and 2002 (Sensormedics model 2200, USA) and by applying stringent quality control criteria [43]. The forced expiratory manoeuvre was obtained without bronchodilators. FEV1 and FVC had to originate from the same manoeuvre in order to provide a valid FEV1/FVC ratio. Information about the smoking history was collected by questionnaire. Passive smoking was positive if never smoking subjects gave an affirmative answer at baseline or follow-up to the question if they were exposed to environmental tobacco smoke in the 12 months prior to the examination on most days or nights. Height and weight were measured and BMI was calculated as weight divided by squared height. Incident cases of airflow obstruction were defined as persons with a FEV1/FVC ratio $\geq$ 0.7 at baseline and < 0.7 at follow-up and were compared to individuals without obstruction at both examinations. Incident cases of respiratory symptoms were defined as people with self-reported regular cough, phlegm or shortness of breath at follow-up, but not at baseline. They were compared with individuals without any of these symptoms at baseline and follow-up. Cough or phlegm had to be present during the day or at night on most days for as much as 3 months per year and shortness of breath had to occur during sleep in the past 12 months before the examination. Subjects who declared an asthma diagnosis by a physician at baseline or follow-up were defined as asthmatics.

Serum Analysis

AAT and hs-CRP concentrations were determined from blood serum aliquots by latex-enhanced immunoturbidimetric assays (Roche diagnostics, Germany). Lower detection thresholds for the AAT and CRP assays were 0.21 g/l and 0.1 mg/l, respectively.

Genotyping

Genotyping of *SERPINA1* PiS (rs17580) and PiZ (rs28929474) polymorphisms was carried out using 5′ nuclease fluorescent real-time PCR (TaqMan Probes technology) on LightCycler480 (Roche) as described before [4]. Probes and primers are given in Table S4. Genotype distributions for PiS and PiZ were both in Hardy Weinberg equilibrium (p = 0.93, N = 6050, and p = 0.99, N = 6051, respectively).

Statistical Analysis

Statistical tests to evaluate differences in the characteristics among the different groups of genotype carriers encompassed Pearson's χ^2 for testing equal proportions, analysis of variance (ANOVA) for testing equal means of normally distributed continuous data, and Kruskal-Wallis for testing equal distributions of continuous data which were not normally distributed. Main effects of *SERPINA1* alleles on lung function decline were assessed using multiple unconditional linear regression models adjusted for sex, age, recruiting area, smoking history (packyears at baseline and between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up. Age and packyears between baseline and follow-up were modeled with linear and squared terms to better fit to spirometry data. Interactions between genotypes and other covariates were tested by integrating

multiplicative terms in the regression models. Two-sided p-values of <0.05 (and of <0.10 for interactions) were considered as statistically significant. We performed 56 different linear regression tests (4 respiratory outcomes * 2 genotype comparisons * 7 categories). Bonferroni correction would thus lower the significance threshold to $p = 0.05/56 = 0.001$. However, since all analyses were hypothesis-driven and most tests not independent of each other, we decided to give the results uncorrected for multiple testing.

Logistic regression models were used to compare the odds of developing airflow obstruction or respiratory symptoms between baseline and follow-up among the *SERPINA1* genotype classes. The models were adjusted for the same covariates mentioned above. All statistical analyses were performed with STATA, release 10.1 IC (STATA corporation, USA).

Supporting Information

Table S1 Sensitivity analyses for adjusted mean values in Δ(FEV1/FVC) and ΔFEF25-75% over 11 years of follow-up comparing different *SERPINA1* genotypes.
(PDF)

Table S2 Adjusted mean values in Δ (FEF25-75%/FVC) over 11 years of follow-up comparing different *SERPINA1* genotypes.
(PDF)

Table S3 Adjusted mean values in lung function change over 11 years of follow-up comparing different *SERPINA1* genotypes in unweighted and weighted models.
(PDF)

Table S4 Primers and probes for genotyping the *SERPINA1* PiS and PiZ polymorphisms (rs17580 and rs28929474) using 5′ nuclease fluorescent real-time PCR (TaqMan Probes technology) on LightCycler480 (Roche).
(PDF)

Author Contributions

Conceived and designed the experiments: TR EWR ML NPH. Performed the experiments: GAT IF MI MZ SO. Analyzed the data: GAT IF MI TR NPH. Contributed reagents/materials/analysis tools: MG FK POB EZ. Wrote the paper: GAT NPH. Contributed to the interpretation of the data: IF MI TR NPH .

References

1. Hall IP, Lomas DA The genetics of obstructive lung disease: big is beautiful. Thorax (2010) 65: 760-761.
2. Janus ED, Phillips NT, Carrell RW Smoking, lung function, and alpha 1-antitrypsin deficiency. Lancet (1985) 1: 152-154.
3. Luisetti M, Seersholm N Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax (2004) 59: 164-169.
4. Ferrarotti I, Thun GA, Zorzetto M, Ottaviani S, Imboden M, et al. Serum levels and genotype distribution of alpha1-antitrypsin in the general population. Thorax(2012).
5. Dahl M, Hersh CP, Ly NP, Berkey CS, Silverman EK, et al. The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J (2005) 26: 67-76.
6. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, et al. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax (2004) 59: 843-849.
7. Dahl M, Tybjaerg-Hansen A, Lange P, Vestbo J, Nordestgaard BG Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha1-antitrypsin MZ heterozygotes: A longitudinal study of the general population. Ann Intern Med (2002) 136: 270-279.
8. Wadsworth ME, Vinall LE, Jones AL, Hardy RJ, Whitehouse DB, et al. Alpha1-antitrypsin as a risk for infant and adult respiratory outcomes in a national birth cohort. Am J Respir Cell Mol Biol (2004) 31: 559-564.
9. Silva GE, Sherrill DL, Guerra S, Barbee RA A longitudinal study of alpha1-antitrypsin phenotypes and decline in FEV1 in a community population. Chest (2003) 123: 1435-1440.
10. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, et al.Susceptibility genes for rapid decline of lung function in the lung health study. Am J Respir Crit Care Med (2001) 163: 469-473.
11. Johnson D, Travis J The oxidative inactivation of human alpha-1-proteinase inhibitor. Further evidence for methionine at the reactive center. J Biol Chem (1979) 254: 4022-4026.
12. Alam S, Li Z, Janciauskiene S, Mahadeva R Oxidation of Z alpha1-antitrypsin by cigarette smoke induces polymerization: a novel mechanism of early-onset emphysema. Am J Respir Cell Mol Biol (2011) 45: 261-269.
13. Gan WQ, Man SF, Senthilselvan A, Sin DD Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis. Thorax (2004) 59: 574-580.
14. McKeever T, Saha S, Fogarty AW The association between systemic inflammatory cellular levels and lung function: a population-based study. PLoS One (2011) 6: e21593.
15. Senn O, Russi EW, Schindler C, Imboden M, von Eckardstein A, et al. Circulating alpha1-antitrypsin in the general population: determinants and association with lung function. Respir Res (2008) 9: 35.
16. Vonk JM, Jongepier H, Panhuysen CI, Schouten JP, Bleecker ER, et al. Risk factors associated with the presence of irreversible airflow limitation and reduced transfer coefficient in patients with asthma after 26 years of follow up. Thorax (2003)58: 322-327.
17. Eriksson S, Lindell SE, Wiberg R Effects of smoking and intermediate alpha 1-antitrypsin deficiency (PiMZ) on lung function. Eur J Respir Dis (1985) 67: 279-285.
18. Senn O, Russi EW, Imboden M, Probst-Hensch NM alpha1-Antitrypsin deficiency and lung disease: risk modification by occupational and environmental inhalants. Eur Respir J (2005) 26: 909-917.
19. von Ehrenstein OS, von Mutius E, Maier E, Hirsch T, Carr D, et al. Lung function of school children with low levels of alpha1-antitrypsin and tobacco smoke exposure. Eur Respir J (2002) 19: 1099-1106.
20. Banauch GI, Brantly M, Izbicki G, Hall C, Shanske A, et al. Accelerated spirometric decline in New York City firefighters with alpha-antitrypsin deficiency. Chest (2010) 138: 1116-1124.
21. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet (2012) 21: 947-957.

22. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet (2011) 43: 1082-1090.

23. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, et al. Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol (2012) 129: 1218-1228.

24. Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, et al. A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS One (2011)6: e19382.

25. Mulgrew AT, Taggart CC, Lawless MW, Greene CM, Brantly ML, et al. Z alpha1-antitrypsin polymerizes in the lung and acts as a neutrophil chemoat-tractant. Chest (2004) 125: 1952-1957.

26. Malerba M, Ricciardolo F, Radaeli A, Torregiani C, Ceriani L, et al. Neutrophilic inflammation and IL-8 levels in induced sputum of alpha-1-antitrypsin PiMZ subjects. Thorax (2006) 61: 129-133.

27. Salome CM, King GG, Berend N Physiology of obesity and effects on lung function. J Appl Physiol (2010) 108: 206-211.

28. Roytblat L, Rachinsky M, Fisher A, Greemberg L, Shapira Y, et al. Raised interleukin-6 levels in obese patients. Obes Res (2000) 8: 673-675.

29. Bullo M, Garcia-Lorda P, Megias I, Salas-Salvado J Systemic inflammation, adipose tissue tumor necrosis factor, and leptin expression. Obes Res (2003) 11: 525-531.

30. Sunyer J, Pistelli R, Plana E, Andreani M, Baldari F, et al. Systemic inflammation, genetic susceptibility and lung function. Eur Respir J (2008) 32: 92-97.

31. Margretardottir OB, Thorleifsson SJ, Gudmundsson G, Olafsson I, Benedikts-dottir B, et al. Hypertension, systemic inflammation and body weight in relation to lung function impairment-an epidemiological study. COPD (2009) 6: 250-255.

32. Bridevaux PO, Gerbase MW, Schindler C, Dietrich DF, Curjuric I, et al. Sex-specific effect of body weight gain on systemic inflammation in subjects with COPD: results from the SAPALDIA cohort study 2. Eur Respir J (2009) 34: 332-339.

33. Dahl M, Vestbo J, Zacho J, Lange P, Tybjaerg-Hansen A, et al. C reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach. Thorax (2011) 66: 197-204.

34. Cochrane GM, Prieto F, Hickey B, Benatar SR, Clark TJ Early diagnosis of airways obstruction. Thorax (1974) 29: 389-393.

35. Simon MR, Chinchilli VM, Phillips BR, Sorkness CA, Lemanske RF Jr, et al. Forced expiratory flow between 25% and 75% of vital capacity and FEV1/forced vital capacity ratio in relation to clinical and physiological parameters in asthmatic children with normal FEV1 values. J Allergy Clin Immunol (2010) 126:527-534 e521-528.

36. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, et al. alpha-Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. Chest (2010) 138: 1125-1132.

37. DeMeo DL, Carey VJ, Chapman HA, Reilly JJ, Ginns LC, et al. Familial aggregation of FEF(25-75) and FEF(25-75)/FVC in families with severe, early onset COPD. Thorax (2004) 59: 396-400.

38. Kabesch M, Hoefler C, Carr D, Leupold W, Weiland SK, et al. Glutathione S transferase deficiency and passive smoking increase childhood asthma. Thorax (2004) 59: 569-573.

39. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al. Standardisation of spirometry. Eur Respir J (2005) 26: 319-338.

40. Downs SH, Schindler C, Liu LJ, Keidel D, Bayer-Oglesby L, et al. Reduced exposure to PM10 and attenuated age-related decline in lung function. N Engl J Med (2007) 357: 2338-2347.

41. Imboden M, Schwartz J, Schindler C, Curjuric I, Berger W, et al. Decreased PM10 exposure attenuates age-related lung function decline: genetic variants in p53, p21, and CCND1 modify this effect. Environ Health Perspect (2009) 117: 1420-1427.

42. Curjuric I, Imboden M, Schindler C, Downs SH, Hersberger M, et al. HMOX1 and GST variants modify attenuation of FEF25-75% decline due to PM10 reduction. Eur Respir J (2010) 35: 505-514.

43. Kunzli N, Kuna-Dibbert B, Keidel D, Keller R, Brandli O, et al. Longitudinal validity of spirometers–a challenge in longitudinal studies. Swiss Med Wkly (2005) 135: 503-508.

44. Probst-Hensch NM, Curjuric I, Pierre-Olivier B, Ackermann-Liebrich U, Bettschart RW, et al. Longitudinal change of prebronchodilator spirometric obstruction and health outcomes: results from the SAPALDIA cohort. Thorax (2010) 65: 150-156.

45. Thomason MJ, Strachan DP Which spirometric indices best predict subsequent death from chronic obstructive pulmonary disease? Thorax (2000) 55: 785-788.

46. Martin BW, Ackermann-Liebrich U, Leuenberger P, Kunzli N, Stutz EZ, et al. SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. Soz Praventivmed (1997) 42: 67-84.

47. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, et al. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed (2005) 50: 245-263.

48. Zorzetto M, Russi E, Senn O, Imboden M, Ferrarotti I, et al. SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. Clin Chem (2008) 54: 1331-1338.

Table S1. Sensitivity analyses for adjusted mean values in Δ(FEV1/FVC) and ΔFEF25-75% over 11 years of follow-up comparing different SERPINA1 genotypes.

All	n	ΔFEF25-75%(ml/y)	p-value	Δ(FEV/FVC) (‰/y)	p-value
MM	4207	-70.60		-4.03	
MS	356	-74.42	0.26	-4.14	0.65
MZ	112	-81.41	0.07	-4.46	0.33
MM, age>30	3320	-70.93		-3.84	
MS, age>30	291	-75.20	0.25	-3.97	0.67
MZ, age>30	90	-84.42	0.04	-4.47	0.21
MM, non-asthmatics	3751	-71.41		-3.99	
MS, non-asthmatics	316	-73.17	0.63	-3.95	0.90
MZ, non-asthmatics	102	-80.82	0.13	-4.39	0.38
Ever smokers					
MM	2194	-70.02		-3.84	
MS	179	-74.91	0.32	-4.16	0.38
MZ	52	-87.44	0.05	-4.80	0.14
MM, age>30	1800	-69.12		-3.78	
MS, age>30	156	-73.48	0.41	-3.92	0.71
MZ, age>30	44	-86.62	0.07	-4.65	0.22
MM, non-asthmatics	1957	-70.52		-3.78	
MS, non-asthmatics	156	-71.52	0.85	-3.84	0.86
MZ, non-asthmatics	46	-86.07	0.10	-4.74	0.16
Persistent smokers					
MM	922	-66.82		-3.90	
MS	74	-74.02	0.34	-4.62	0.21
MZ	18	-108.16	0.005	-5.25	0.23
MM, age>30	708	-66.63		-4.03	
MS, age>30	62	-75.62	0.27	-4.61	0.37

MZ, age>30	12	-105.47	0.03	-4.54	0.72
MM, non-asthmatics	844	-66.81		-3.77	
MS, non-asthmatics	64	-68.99	0.79	-4.08	0.62
MZ, non-asthmatics	18	-107.33	0.007	-5.09	0.24
Obese Subjects					
MM	653	-58.39		-2.81	
MS	55	-59.46	0.90	-3.61	0.22
MZ	16	-92.16	0.03	-5.13	0.05
MM, age>30	581	-68.14		-3.19	
MS, age>30	51	-64.06	0.64	-3.56	0.59
MZ, age>30	16	-99.22	0.04	-5.49	0.05
MM, non-asthmatics	561	-59.56		-2.75	
MS, non-asthmatics	47	-58.67	0.92	-3.38	0.36
MZ, non-asthmatics	14	-95.23	0.03	-5.32	0.04
Subjects in upper tertile of hs-CRP					
MM	1387	-71.24		-4.00	
MS	99	-89.85	0.003	-5.03	0.04
MZ	36	-99.32	0.006	-5.46	0.07
MM, age>30	1174	-70.80		-3.98	
MS, age>30	86	-89.02	0.006	-5.00	0.06
MZ, age>30	31	-94.91	0.03	-5.54	0.08
MM, non-asthmatics	1229	-71.47		-3.86	
MS, non-asthmatics	86	-87.06	0.02	-4.55	0.20
MZ, non-asthmatics	33	-101.16	0.006	-5.62	0.04

Covariates included sex, linear and squared age, recruiting area, smoking history (packyears at baseline, as well as linear and squared packyears between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up.

Persistent smokers were classified as subjects who declared current smoking at both examinations.

Obese subjects were defined as BMI $\geq 30 \text{kg/m}^2$ at the baseline or follow-up examination.

Subjects in the upper tertile of hs-CRP had blood levels of ≥ 1.8 mg/l.

Table S2. Adjusted mean values in Δ(FEF25-75%/FVC) over 11 years of follow-up comparing different SERPINA1 genotypes.

All	Δ(FEF25-75%/FVC) (‰/y)	95%CI	p-value
MM, n=4207	-13.09	-13.56 to -12.62	
MS, n=356	-13.65	-15.28 to -12.01	0.52
MZ, n=112	-15.80	-18.70 to -12.89	0.07
Ever smokers			
MM, n=2194	-12.77	-13.51 to -12.04	
MS, n=179	-13.75	-16.05 to -11.44	0.42
MZ, n=52	-17.37	-21.59 to -13.15	0.03
Persistent smokers			
MM, n=922	-12.50	-14.57 to -10.43	
MS, n=74	-13.68	-17.55 to -9.82	0.52
MZ, n=18	-21.58	-28.94 to -14.22	0.01
Obese subjects			
MM, n=653	-9.33	-12.45 to -6.21	
MS, n=55	-10.01	-15.16 to -4.87	0.76
MZ, n=16	-17.67	-26.07 to -9.28	0.04
Subjects in upper tertile of hs-CRP			
MM, n=1387	-13.05	-14.02 to -12.07	
MS, n=99	-16.77	-20.07 to -13.47	0.03
MZ, n=36	-20.80	-26.21 to -15.38	0.006

Covariates included sex, linear and squared age, recruiting area, smoking history (packyears at baseline, as well as linear and squared packyears between baseline and follow-up), height, baseline BMI and BMI change between baseline and follow-up.
Persistent smokers were classified as subjects who declared current smoking at both examinations.
Obese subjects were defined as BMI ≥ 30kg/m^2 at the baseline or follow-up examination.
Subjects in the upper tertile of hs-CRP had blood levels of ≥ 1.8 mg/l.

Table S3. Adjusted mean values in lung function change over 11 years of follow-up comparing different SERPINA1 genotypes in unweighted and weighted models.

All	n	ΔFEV1(ml/y)	p-value	ΔFVC(ml/y)	p-value	ΔFEF25-75%(ml/y)	p-value	Δ(FEV/FVC) (‰/y)	p-value
MM	4207	-35.16		-24.11		-70.60		-4.03	
MS	356	-36.91	0.24	-24.68	0.77	-74.42	0.26	-4.14	0.65
MZ	112	-36.02	0.74	-23.27	0.80	-81.41	0.07	-4.46	0.33
MM, weighted	4207	-35.15		-24.14		-70.57		-4.02	
MS, weighted	356	-36.88	0.25	-24.55	0.84	-74.49	0.24	-4.16	0.60
MZ, weighted	112	-36.21	0.68	-23.64	0.89	-81.61	0.10	-4.47	0.29

The weights were calculated inverse to the probability of being included in the study sample (based on age, sex and recruiting area variables).

Table S4. Primers and probes for genotyping the SERPINA1 PiS and PiZ polymorphisms (rs17580 and rs28929474) using 5' nuclease fluorescent real-time PCR (TaqMan Probes technology) on LightCycler480 (Roche).

	PiS (rs17580)	PiZ (rs28929474)
Primer Forward	5'-GCCATCTTCTTCCTGCCTGAT-3'	5'-TCCAAGGCCGTGCATAAGG-3'
Primer Reverse	5'-CCAGGAACTTGGTGATGATATCGT-3'	5'-GCCCCAGCAGCTTCAGT-3'
VIC dye-labeled probe	5'-CACCTGGAAAATGAA-3'	5'-ACCATCGACGAGAAAG-3'
FAM dye-labeled probe	5'-CACCTGGTAAATGAA-3'	5'-CATCGACAAGAAAG-3'

5.4 Paper 5: Interactions between *SERPINA1* PiMZ Genotype, Occupational Exposure, and Lung Function Decline.

This paper was published:

Mehta AJ*, **Thun GA***, Imboden M, Ferrarotti I, Keidel D, Künzli N, Kromhout H, Miedinger D, Phuleria H, Rochat T, Russi EW, Schindler C, Schwartz J, Vermeulen R, Luisetti M, Probst-Hensch NM. Occupational and Environmental Medicine. Published Online First: 08 November 2013.

(*contributed equally to the manuscript)

ORIGINAL ARTICLE

Interactions between *SERPINA1* PiMZ genotype, occupational exposure and lung function decline

A J Mehta,[1,2,3] G A Thun,[1,2] M Imboden,[1,2] I Ferrarotti,[4] D Keidel,[1,2] N Künzli,[1,2]
H Kromhout,[5] D Miedinger,[6,7] H Phuleria,[1,2] T Rochat,[8] E W Russi,[9] C Schindler,[1,2]
J Schwartz,[3] R Vermeulen,[5] M Luisetti,[4] N Probst-Hensch,[1,2] SAPALDIA team[1,2]

ABSTRACT

Objectives We evaluated interactions between *SERPINA1* PiMZ genotype, associated with intermediate α1-antitrysin deficiency, with outdoor particulate matter ≤10 μm (PM_{10}), and occupational exposure to vapours, dusts, gases and fumes (VGDF), and their effects on annual change in lung function.

Methods Pre-bronchodilator spirometry was performed in 3739 adults of the Swiss Cohort Study on Air Pollution and Lung Disease in Adults (SAPALDIA) for whom *SERPINA1* genotypes were available. At baseline in 1991, participants were aged 18–62 years; follow-up measurements were conducted from 2001 to 2003. In linear mixed regression models of annual change in lung function, multiplicative interactions were evaluated between PiMZ genotype (PiMM as reference) and change in PM_{10} ($\mu g/m^3$), and VGDF exposure (high-level, low-level or no exposure as reference) during follow-up.

Results Annual declines in forced expiratory flow at 25–75% of forced vital capacity ($FEF_{25-75\%}$) (−82 mL/s, 95% CI −125 to −39) and forced expiratory volume in 1 s over forced vital capacity (FEV_1/FVC) (−0.3%, 95% CI −0.6% to 0.0%) in association with VGDF exposure were observed only in PiMZ carriers ($P_{interaction} < 0.0001$ and $P_{interaction} = 0.03$, respectively). A three-way interaction between PiMZ genotype, smoking and VGDF exposure was identified such that VGDF-associated $FEF_{25-75\%}$ decline was observed only in ever smoking PiMZ carriers ($P_{interaction} = 0.01$). No interactions were identified between PiMZ genotype and outdoor PM_{10}.

Conclusions *SERPINA1* PiMZ genotype, in combination with smoking, modified the association between occupational VGDF exposure and longitudinal change in lung function, suggesting that interactions between these factors are relevant for lung function decline. These novel findings warrant replication in larger studies.

What this paper adds

► There is evidence that α1-antitrypsin (AAT) deficiency, occupational exposure to vapours, gas, dusts and fumes (VGDF), ambient air pollution and smoking are all causative factors for adverse pulmonary health. The interactions between these factors are not well investigated.

► In the present study, a three-way interaction was identified between *SERPINA1* PiMZ genotype, which is associated with intermediate AAT deficiency, smoking and VGDF exposure. No interactions were present between PiMZ genotype and air pollution.

► Genetic predisposition for intermediate AAT deficiency, in combination with smoking, may increase susceptibility to occupational exposure-related decline in lung function. These findings should be reassessed in larger studies.

INTRODUCTION

Decline in lung function over time is a characteristic of aging, and accelerated decline is distinctive for obstructive lung diseases including asthma and chronic obstructive pulmonary disease (COPD).[1 2] Epidemiological evidence from occupational and population-based studies indicates that occupational exposures to dusts, gases, and fumes are associated with additional decline in lung function and a causal risk factor of COPD.[3 4] There is limited evidence to support that ambient air pollution is associated with lung function decline and COPD.[4–6]

The best known genetic risk factor for COPD is severe serum deficiency of α1-antitrypsin (AAT), an anti-inflammatory protease inhibitor (PI) of neutrophil elastase, a protease which breaks down elastin in the lung parenchyma.[4 7] AAT is encoded by the *SERPINA1* gene and allelic variants of *SERPINA1* lead to inherited low AAT serum levels. Severe AAT deficiency is caused by rare *SERPINA1* genotypes occurring at a frequency of <0.1% in the general population, like homozygosity for the PI deficient Z allele (PiZZ) or compound heterozygosity for PI deficient S and Z alleles (PiSZ).[8] PiMZ and PiMS genotypes of *SERPINA1* (M being the common non-deficient allele) lead to intermediate and mild AAT deficiency, respectively, and are more prevalent in the general populations representative of Middle European countries at 2% and 8% frequency, respectively.[9] Meta-analyses offer mixed evidence to support an association between PiMZ genotype and risk of COPD, independent of smoking; there is no evidence of an association between PiMS genotype and risk of COPD.[10 11]

Earlier studies from the follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) identified associations between improvements in the concentrations of outdoor particulate matter of 10 μm or less (PM_{10}) and attenuation of lung function decline, with the

greatest change in small airway function,[5] and between high levels of occupational exposure to vapours, dusts, gases and fumes (VGDF) and incidence of COPD.[12] A more recent study from the same survey showed associations between PiMZ genotype and age-related decline in forced expiratory flow between 25% and 75% of the forced vital capacity ($\Delta FEF_{25-75\%}$) in participants with systemic low-grade inflammation, including persistent smokers.[13] For this analysis, we evaluated whether the PiMZ genotype modified the associations between environmental and occupational exposures, including outdoor PM_{10}, environmental tobacco smoke (ETS) exposure and occupational exposure to VGDF, and annual change in lung function in SAPALDIA, hypothesising that the associations between these exposures and lung function decline would be stronger in PiMZ carriers in comparison with PiMM carriers.

METHODS
More detailed descriptions of the SAPALDIA study population and design are summarised in previous publications[12–14] and in the online supplement.

Study population
SAPALDIA is a multicentre, population-based prospective cohort study consisting of a random sample of 9651 participants aged between 18 and 62 years that were recruited from eight regions in Switzerland, and were administered medical examinations, pre-bronchodilator spirometry testing and a detailed health questionnaire, at baseline in 1991.[15] The follow-up survey (SAPALDIA 2) was conducted from 2001 to 2003, of which 8047 of the original study participants were present and of whom 6058 subjects provided blood samples and consented to DNA analysis. This analysis was restricted to 3739 participants who were either SERPINA1 PiMZ or PiMM carriers, completed spirometry testing at both surveys, reported information on important covariates, and for whom air pollution measurements (outdoor PM_{10}) and occupational history during follow-up were available. Because we identified no presence of interaction between PiMS genotype and smoking in our previous analysis,[13] and because there is no evidence of an association between PiMS genotype and risk of COPD, SAPALDIA participants who were PiMS carriers were not included in this analysis. See online supplementary figure E1 for additional details describing the selection of participants included in this analysis. The SAPALDIA cohort study complies with the Declaration of Helsinki. Written informed consent was obtained from participants in both surveys. The study was approved by the central ethics committee of the Swiss Academy of Medical Sciences and the respective Cantonal ethics committees of the eight study regions.

Spirometry testing
The spirometry protocol, which complied with American Thoracic Society (ATS) criteria,[16] has been described in detail elsewhere.[17–19] No bronchodilation was applied. The outcomes of interest for this study were annual changes in forced expiratory volume in 1 s (FEV_1, mL), forced vital capacity (FVC, mL), the ratio of FEV_1 over FVC (FEV_1/FVC, %), and forced expiratory flow between 25% and 75% of FVC (FEF_{25-75}, mL/s). Annual change in lung function was defined as the difference in each variable between the two examinations, divided by the follow-up time in years for the participant. See the online supplement for additional details describing methods for spirometry testing.

Occupational and environmental exposure assessment
ETS exposure at baseline and follow-up visits was assessed for different environments by the question 'How many hours per day are you exposed to other people's tobacco smoke: (i) at home; (ii) at the workplace?' ETS exposure was categorised into three exposure groups: none, <3 h/day but not none, and ≥3 h/day.

Current job title reported by participants at the baseline survey (1991) and all job titles reported during the follow-up period were standardised according to the International Standard Classification of Occupations (ISCO-88) code's four-digit classification,[20] and linked to the ALOHA GPJEM for COPD.[21 22] Evaluation of occupational exposure in this analysis was restricted to exposure to high VGDF level during follow-up (low level or no exposure as reference) and continuous cumulative exposure (>0 years for exposed, 0 years for unexposed; unit, years) during follow-up. See the online supplement for additional descriptions of the ALOHA GPJEM and exposure variables.

Outdoor PM_{10} concentrations ($\mu g/m^3$) outside of each subject's residence were estimated for the years 1990 and 2000 using a validated dispersion model (with different emissions inventories for both years) with a spatial resolution of 200×200 m. We estimated the change in PM_{10} exposure during follow-up (ΔPM_{10}) for each participant, which was the difference between home outdoor mean PM_{10} level averaged over 1 year prior to the follow-up survey (in 2001–2003) and the corresponding mean PM_{10} level averaged over 1 year prior to the baseline survey (in 1990), divided by the time between baseline and follow-up examinations. As documented previously, overall exposure to home outdoor PM_{10} declined over the follow-up period (median, $-5.3 \mu g/m^3$). Further details of the air pollution exposure assessment and modelling have been described previously.[23] For this analysis, ΔPM_{10} was scaled per 10 µg increase over 10 years.

Genotyping
Genomic DNA was extracted from blood samples using the Puregene DNA Isolation Kit (Gentra Systems, USA). DNA samples of 6058 probands were shipped to the Center for Diagnosis of Inherited Alpha1-antitrypsin Deficiency in Pavia (Italy) and processed in a semi-automated medium throughput setup, assisted by liquid handling station (Freedom EVO75, Tecan Group, Switzerland). The PiS (rs17580) and PiZ (rs28929474) single nucleotide polymorphisms (SNPs) of SERPINA1 were genotyped using 5′-nuclease fluorescent real-time PCR (TaqMan Probes technology) on LightCycler480 (Roche). Genotype distributions for PiS and PiZ in this study sample were all in Hardy–Weinberg equilibrium.

Statistical analysis
We used linear mixed regression models of annual changes in FEV_1, FVC, FEV_1/FVC and $FEF_{25-75\%}$, with random intercept for study area, to evaluate multiplicative interactions between SERPINA1 PiMZ genotype (PiMM genotype as reference) and daily ETS, ΔPM_{10} and high-level VGDF exposures during follow-up. Interactions between PiMZ genotype and each exposure type were evaluated simultaneously in all models. We extracted the β coefficients to estimate the adjusted difference in annual change in lung function in association with each exposure in PiMZ and PiMM carriers from a single nested model with the multiplicative interaction terms. All models were adjusted for exposure and covariates ascertained at baseline

including PM_{10} concentration, cumulative VGDF exposure, age, age squared, sex, height (cm), body mass index (BMI, kg/m^2), early respiratory infection, parental asthma, and education; additional covariates included smoking status through follow-up (persistent smokers, former smokers, never smokers as reference), cumulative pack-years through follow-up, the interaction PiMZ genotype and smoking status, difference in BMI over follow-up period, and PiMZ genotype. Additionally, we tested for three-way interaction between smoking status through follow-up (ever smokers, never smokers), PiMZ genotype, and either ΔPM_{10} or high-level VGDF exposure. All analyses were done using SAS V.9.2. Two-sided p values <0.05 were interpreted as statistically significant for main and interaction effects.

Secondary analyses

A substantial number of participants with complete information at baseline were not included in the analysis due to non-participation, or incomplete genotyping, spirometry, or exposure and other covariate information at follow-up (49.5%). The non-participants were more likely to be older, female, heavier, smokers, have lower lung function, report more daily exposure to ETS, or have higher exposure to outdoor PM_{10} at the baseline survey (see online supplementary table E1). No significant difference was observed in the distribution of occupational VGDF exposure between the participants and non-participants. To account for the potential bias arising from differences between participants and non-participants, we weighted the linear mixed effect regression models by the inverse probability for inclusion in the analysis. We also evaluated whether baseline lung function attenuated the interactions between PiMZ genotype and each exposure, with additional adjustment for baseline lung function.

RESULTS

Characteristics of the study population stratified by *SERPINA1* genotype are summarised in table 1. Similar distributions between genotypes were observed for BMI and height, but PiMZ carriers were slightly younger, and had higher lung function at baseline on average compared with PiMM carriers. The mean annual declines in FEV_1 and FEV_1/FVC over the follow-up period were slightly larger in PiMZ carriers compared with PiMM carriers, with a larger difference observed for $FEF_{25-75\%}$ (−88 mL/s vs −71 mL/s, respectively). The prevalence of ever smokers and mean cumulative pack-years was lower in PiMZ carriers. Compared with PiMM carriers, fewer PiMZ carriers reported being exposed to ETS ≥3 h/day. Over the follow-up period, PiMZ carriers had a slightly larger mean reduction in outdoor PM_{10} on average compared with PiMM carriers. The prevalence of high-level VGDF exposure at baseline and during follow-up was also moderately lower in PiMZ carriers.

The adjusted differences in annual changes in lung function in association with ΔPM_{10} and high-level VGDF exposures during follow-up estimated in PiMZ and PiMM carriers are presented in table 2. A statistically significant interaction (p<0.0001) was observed between PiMZ genotype and high-level VGDF on annual change in $FEF_{25-75\%}$. High-level VGDF exposure was associated with accelerated annual decline in $FEF_{25-75\%}$ (−82 mL/s, 95% CI −125 to −39) in PiMZ carriers, while a positive association was observed in PiMM carriers. A similar interaction of statistical significance (p=0.03) was observed between PiMZ genotype and high-level VGDF exposure on annual change in FEV_1/FVC. Overall, larger annual declines in lung function in association with ΔPM_{10} were

observed in PiMZ carriers than in PiMM carriers, but no statistically significant interactions were present. No interactions were present between PiMZ genotype and daily ETS exposure (data not shown).

After adjustment for baseline lung function (see online supplementary table E2), the estimated effects of ΔPM_{10} and high-level VGDF exposure on annual decline in lung function in PiMZ carriers were smaller compared to those presented in table 2, but the interaction between PiMZ genotype and high-level VGDF exposure on annual change in $FEF_{25-75\%}$ remained statistically significant. The estimated associations presented in table 2 were similar to those after weighting the models according to the inverse probability of inclusion in the analysis (see online supplementary table E3); the interaction between PiMZ genotype and high-level VGDF exposure on annual change in FEV_1/FVC and $FEF_{25-75\%}$ remained statistically significant. When occupational VGDF exposure was characterised as continuous cumulative exposure (see online supplementary table E4), the interactions between PiMZ genotype and cumulative VGDF exposure during follow-up were similar to those presented for high-level VGDF exposure presented in table 2; statistically significant excess annual declines in FEV_1/FVC and $FEF_{25-75\%}$ per 10 years of cumulative VGDF exposure were observed only in PiMZ carriers.

Table 3 summarises the adjusted differences in annual change in lung function for ΔPM_{10} and high-level VGDF exposure by PiMZ genotype and smoking status, as estimated from the nested model with a three-way interaction between PiMZ genotype, smoking, and environmental (ΔPM_{10}) or occupational (high-level VGDF) exposure, respectively. Overall, the largest annual declines in $FEF_{25-75\%}$ (−109 mL/s; 95% CI −159 to −59) and FEV_1/FVC (−0.4%; 95% CI −0.8% to −0.0%) associated with high-level VGDF exposure during follow-up were observed for ever smoking PiMZ carriers. A three-way interaction between ever smoking status, PiMZ genotype and high-level VGDF exposure on annual change in $FEF_{25-75\%}$ was statistically significant (p=0.01).

DISCUSSION

In this prospective cohort study of Swiss working adults, we observed the *SERPINA1* PiMZ genotype, in combination with smoking, to modify the association between occupational exposure to VGDF during follow-up and annual change in lung function, particularly for $FEF_{25-75\%}$ and FEV_1/FVC such that the excess annual decline in $FEF_{25-75\%}$ and FEV_1/FVC was observed in ever smoking PiMZ carriers. To a lesser extent, larger annual declines in lung function in association with ΔPM_{10} were observed in PiMZ carriers than in PiMM carriers, but no statistically significant interactions were present. We did not detect any interaction between PiMZ genotype and daily ETS exposure.

Earlier cross-sectional population-based studies of homozygous PiZ individuals have shown that lower FEV_1 was observed in participants who reported occupational exposures compared to unexposed participants.[24][25] Similarly, a prospective study of exposed iron-ore miners demonstrated that a larger decline in FEV_1 was observed in workers with intermediate AAT deficiency compared with workers with normal AAT phenotype.[26] These studies could not formally assess effect modification as they included only AAT deficient or only exposed individuals. A cross-sectional study of young farming apprentices and unexposed conscripts observed an interaction between farming occupation and PiMZ phenotype on bronchial hyperresponsiveness; no effect modification by PiMZ phenotype was observed between farming occupation and lung function.[27] A recent

Table 1 Distribution of characteristics for PiMM and PiMZ carriers

SERPINA 1 genotype	PiMM	PiMZ
Number of participants	n=3642	n=97
Lung function at baseline		
FEV$_1$, L, mean (SD)	3.64 (0.81)	3.70 (0.79)
FVC, L, mean (SD)	4.61 (1.02)	4.60 (0.97)
FEV$_1$/FVC, %, mean (SD)	79.2 (7.5)	81.0 (8.0)
FEF$_{25-75}$, L/s, mean (SD)	3.49 (1.21)	3.81 (1.31)
Annual change in lung function from baseline to follow-up		
$\triangle$FEV$_1$, mL, mean (SD)	−35 (31)	−37 (26)
$\triangle$FVC, mL, mean (SD)	−24 (41)	−23 (42)
$\triangle$FEV$_1$/FVC, %, mean (SD)	−0.4 (0.5)	−0.5 (0.5)
$\triangle$FEF$_{25-75}$, mL/s, mean (SD)	−71 (65)	−88 (75)
Covariates of interest at baseline		
Age, years, mean (SD)	40.4 (11.2)	39.5 (10.9)
BMI, kg/m^2, mean (SD)	23.7 (3.6)	23.5 (3.4)
Height, cm, mean (SD)	169.9 (8.8)	169.6 (9.5)
Male, n (%)	1845 (50.6)	47 (48.5)
Early respiratory infection, n (%)	321 (8.1)	6 (6.2)
Parental history of asthma, n (%)	381 (10.5)	13 (13.4)
Education level, n (%)		
High	650 (17.9)	25 (25.8)
Medium	2585 (71.0)	67 (69.1)
Low	407 (11.2)	5 (5.2)
ETS and smoking status at follow-up		
Daily ETS exposure in past 12 months prior to follow-up, n (%)		
≥3 h/day	355 (9.8)	4 (4.1)
<3 h/day	560 (15.4)	15 (15.5)
Unexposed	2727 (74.9)	78 (80.4)
Smoking status through follow-up, n (%)		
Persistent smokers	877 (24.1)	20 (20.6)
Former smokers	1178 (32.3)	28 (28.9)
Never smokers	1587 (43.6)	49 (50.5)
Cumulative pack-years through follow-up, mean (SD)	10.7 (18.3)	9.7 (15.6)
Occupational VGDF exposure		
VGDF level in current occupation at baseline survey, n (%)		
High	479 (13.2)	7 (7.2)
Low	1190 (32.7)	29 (29.9)
Unexposed	1973 (54.2)	61 (62.9)
Cumulative VGDF exposure at baseline survey among exposed, years, mean (SD)	13.3 (28.0)	10.5 (25.3)
Highest VGDF level in all occupations during follow-up, n (%)		
High	490 (13.5)	9 (9.3)
Low	1178 (32.3)	34 (35.1)
Unexposed	1974 (54.2)	54 (55.7)
Cumulative VGDF exposure during follow-up among exposed, years, mean (SD)	7.2 (12.4)	5.4 (9.6)
Outdoor PM$_{10}$ exposure		
1-year average PM$_{10}$ prior to baseline survey, μg/m^3, mean (SD)	27.0 (9.8)	27.0 (10.1)
Change in PM$_{10}$ between follow-up and baseline surveys, μg/m^3, mean (SD)	−6.2 (4.2)	−6.5 (5.0)

ETS, environmental tobacco smoke; FVC, forced vital capacity; FEF, forced expiratory flow; FEV, forced expiratory volume; VGDF, vapours, dusts, gases and fumes.

study of New York City World Trade Center (WTC) firefighters observed that workers with moderate and mild AAT deficiency experienced larger declines in FEV$_1$ in comparison with participants with no AAT deficiency during the first 4 years after September 11th.[28] In the study of WTC firefighters, AAT deficiency also did not enhance the rate of FEV$_1$ decline among the study participants pre-9/11, suggesting a novel interaction between AAT deficiency and exposures resulting from 9/11 on lung function decline. Our study is the first to identify interaction between *SERPINA1* genotype for predisposition to

intermediate AAT deficiency and chronic occupational exposure to high-levels of VGDF in a general population prospective cohort-based setting over a long duration of follow-up.

It is of interest that PiMZ carriers were less likely to be exposed to high-level VGDF exposures at baseline and during the follow-up period compared with PiMM carriers. The discrepancy in the VGDF exposure distribution by *SERPINA1* genotype may be suggestive of a healthy selection effect associated with occupational inhalant exposures. It may be hypothesised that PiMZ carriers may select themselves away from

Table 2 Adjusted differences* in annual change in lung function in association with outdoor PM_{10} and occupational VGDF exposures during follow-up in PiMZ and PiMM carriers

	N	ΔFEV_1 (mL) β (95% CI)†	ΔFVC (mL) β (95% CI)†	$\Delta FEV_1/FVC$ (%) β (95% CI)†	$\Delta FEF_{25-75\%}$ (mL/s) β (95% CI)†
ΔPM_{10} (per 10 μg/m³ increase in 10 years)					
PiMZ	97	−6 (−15 to 4)	−2 (−18 to 15)	−0.2 (−0.4 to 0.0)	−24 (−52 to 4)
PiMM	3642	−3 (−7 to 0)	−0 (−5 to 5)	−0.0 (−0.1 to 0.0)	−12 (−21 to −3)
High-level VGDF exposure					
PiMZ	97	−8 (−27 to 11)	10 (−16 to 35)	−0.3 (−0.6 to 0.0)‡	−82 (−125 to 39)‡
PiMM	3642	1 (−2 to 4)	−2 (−7 to 2)	0.1 (0.0 to 0.1)	6 (−1 to 13)

*Annual change in lung function was modelled in linear mixed regression with random intercept for area consisting of the following baseline covariates: age, age squared, sex, foreign status, height, body mass index, early respiratory infection, parental asthma, high-level education, cumulative VGDF exposure, and outdoor PM_{10}; additional covariates included smoking status through follow-up, cumulative pack-years through follow-up, daily environmental tobacco smoke exposure at follow-up, difference in body mass index over follow-up period, seasonality, ΔPM_{10}, high-level VGDF exposure during follow-up, and two-way interaction terms between environmental tobacco smoke exposure at follow-up and *SERPINA1* genotype (PiMZ, PiMM as reference), and smoking status through follow-up and *SERPINA1* genotype; we estimated the effect of ΔPM_{10} and high-level VGDF exposure during follow-up in PiMZ and PiMM carriers by adding interaction terms in the model.
†Negative signs indicate acceleration of decline in the lung function parameter with higher levels of exposure to the inhalants.
‡p<0.0001 and p=0.03 for interactions between *SERPINA1* genotype and high-level VGDF exposure on $\Delta FEF_{25-75\%}$ and $\Delta FEV_1/FVC$, respectively.
FVC, forced vital capacity; FEF, forced expiratory flow; FEV, forced expiratory volume; VGDF, vapours, dusts, gases and fumes.

occupational respirable exposures that may be irritable. They may not seek employment in occupations with high exposure to VGDF due to health conditions attributable to AAT deficiency, or they may leave occupations with high exposure to VGDF due to health aggravation or illness prior to the baseline survey, all of which may reduce power to observe an association between occupational exposure and lung function decline in PiMZ carriers. Because we did not have *SERPINA1* genotype information at the time of the baseline survey, we cannot formally evaluate this question, but the influence of genetic susceptibility on healthy worker selection and survival effects merits further research.

There is mixed evidence to support that PiMZ is associated with accelerated lung function decline and increased risk of COPD,[11] and our previous analysis in SAPALDIA did not identify any main associations between PiMZ genotype and lung function decline.[13] However, the current findings suggest that PiMZ genotype, only in combination with smoking, may enhance susceptibility for occupational exposure associated lung function decline. The biological mechanism(s) explaining this susceptibility is not clear. Previous studies in other populations have also observed a combined effect between occupational exposure to VGDF and smoking on risk of COPD.[29–31] Thus the combined interactive effects between PiMZ genotype, smoking and occupational VGDF exposure may be a reflection of smokers with intermediate AAT deficiency exceeding their ability to cope with additional occupational VGDF exposure, or due to poor mucus clearance among PiMZ smokers leading to prolonged inflammatory response after occupational VGDF exposure. Alternatively, it may be that PiMZ individuals may be more susceptible to lung function decline attributable to the combination of VGDF and cigarette smoke due to a diminished protection against tissue damage from elastase in AAT deficiency.[32]

The interaction between causative factors for adverse pulmonary health including AAT deficiency, smoking, and occupational VGDF exposure is also of special public health relevance, particularly because COPD and its related conditions are associated with high morbidity and mortality, and because the prevalence of genetic predisposition to intermediate AAT deficiency is estimated to be around 30 million worldwide.[33] The role of genetic screening in the workplace is also raised as a topic of discussion[34–36] in light of the numerous studies that have documented gene–occupation interactions. However, no genetic test related to an occupational disease has been validated or accepted for use, except the use of genetic biomarkers to measure the dose of a genotoxic exposure[34]; and there are many scientific, ethical, legal and social issues to consider for the use of genetic screening for susceptibility to workplace exposures. In this context, it serves as an important reminder to occupational health practitioners and researchers that occupational exposures are the primary cause of occupational disease and will remain the responsibility of the employer to control.[35]

This study has numerous strengths, including selection of SNPs with demonstrated functionality,[9] prospective study design, detailed data on individual smoking and ETS and PM_{10} exposure, semi-quantitative estimates of occupational exposures to VGDF using a GPJEM, extensive control for confounding, and high quality of longitudinal lung function data. However, there are a number of limitations to be considered. We observed that effect modification by the PiMZ genotype on the association between VGDF exposure and lung function decline was strongest for $FEF_{25-75\%}$, consistent with previous studies of the SAPALDIA cohort that have also identified strong main and interacting associations between air pollution and $FEF_{25-75\%}$,[5 37] and between PiMZ genotype and $FEF_{25-75\%}$ in population subgroups indicative of low-grade systemic inflammation.[13] $FEF_{25-75\%}$ is also used as an indicator of small airways obstruction and bronchial reactivity and correlates highly with percentage predicted FEV_1/FVC at levels of at least moderate airflow obstruction.[38] The primary limitation of $FEF_{25-75\%}$ is that its measurement precision within subjects is lower than of FEV_1.[39] In this study sample, the correlation between baseline and follow-up measurements of FEV_1, $FEF_{25-75\%}$ and FEV_1/FVC were 0.92, 0.81 and 0.72, respectively. While it is unknown why an interaction was not observed between PiMZ and exposures on annual change in FEV_1, a spirometry metric which has less error, we hypothesise, given the previous findings in the SAPALDIA study, that the interactive effects observed for $FEF_{25-75\%}$ are of relevance in this study population.

The relatively high degree of non-participation in our analysis may have also biased our observed associations. However, the use of inverse probability weighting to account for this potential selection bias did not result in considerable differences in the estimated associations, which suggests that any bias from non-participation is likely to be minimal. Finally, VGDF-associated lung function decline was limited to ever smoking PiMZ

Table 3 Adjusted differences* in annual change in lung function during follow-up in association with outdoor PM10 and occupational VGDF exposures during follow-up in PiMZ and PiMM carriers by smoking status

	ΔFEV1 (mL)		ΔFVC (mL)		ΔFEV1/FVC (%)		ΔFEF25-75% (mL/s)	
	Ever smokers β† (95% CI)	Never smokers β† (95% CI)	Ever smokers β† (95% CI)	Never smokers β† (95% CI)	Ever smokers β† (95% CI)	Never smokers β† (95% CI)	Ever smokers β† (95% CI)	Never smokers β† (95% CI)
ΔPM10 (per 10 μg/m³ in 10 years)								
PiMZ (n=48, n=49)‡	−6 (−21 to 10)	−8 (−29 to 12)	1 (−20 to 21)	−9 (−36 to 18)	−0.2 (−0.5 to 0.0)	−0.0 (−0.4 to 0.3)	−28 (−63 to 6)	−9 (−55 to 37)
PiMM (n=2055, n=1587)‡	−2 (−7 to 2)	−5 (−9 to −0)	1 (−5 to 7)	−1 (−7 to 5)	−0.0 (−0.1 to 0.0)	−0.0 (−0.1 to 0.0)	−9 (−17 to 1)	−16 (−26 to −5)
High-level VGDF exposure§								
PiMZ (n=48, n=49)	−4 (−27 to 18)	−18 (−57 to 20)	23 (−5 to 53)	−35 (−86 to 15)	−0.4 (−0.8 to −0.0)	0.1 (−0.5 to 0.8)	−109 (−159 to −59)	9 (−78 to 95)
PiMM (n=2055, n=1587)	4 (−0 to 8)	−2 (−7 to 2)	1 (−5 to 6)	−6 (−12 to −0)	0.1 (−0.0 to 0.1)	0.1 (−0.0 to 0.1)	8 (−0 to 17)	4 (−6 to 13)

*Annual change in lung function was modelled in linear mixed regression with random intercept for area consisting of the following baseline covariates: age, age squared, sex, foreign status, height, body mass index, early respiratory infection, parental asthma, high-level education, cumulative VGDF exposure, and outdoor PM10; additional covariates included smoking status through follow-up, cumulative pack-years through follow-up, daily environmental tobacco smoke exposure at follow-up, difference in body mass index over follow-up period, seasonality, ΔPM10, high-level VGDF exposure during follow-up and SERPINA1 genotype, ΔPM10 and SERPINA1 genotype, and two-way interaction terms between environmental tobacco smoke exposure at follow-up and SERPINA1 genotype, smoking status through follow-up and SERPINA1 genotype, high-level VGDF exposure during follow-up and SERPINA1 genotype, and high-level VGDF exposure during follow-up and smoking status through follow-up; we estimated the effect of ΔPM10 and high-level VGDF exposure during follow-up on annual change in lung function by SERPINA1 genotype and smoking status by adding interaction terms in the model.
†Negative signs indicate acceleration of decline in the lung function parameter with higher levels of exposure to the inhalants.
‡Number of subjects in ever smoker and never smoker subgroups, respectively.
§A statistically significant three-way interaction was observed between SERPINA1 genotype, high-level VGDF exposure and smoking status on ΔFEF25-75% (p<0.0001) and on ΔFEV1/FVC (p=0.02); in ever smokers, statistically significant two-way interactions were observed between SERPINA1 genotype and high-level VGDF exposure on ΔFEF25-75% (p=0.01) and on ΔFEV1/FVC (p=0.02).
FVC, forced vital capacity; FEF, forced expiratory flow; FEV, forced expiratory volume; VGDF, vapours, dusts, gases and fumes.

carriers, which may be explained by insufficient power to identify an interaction between PiMZ genotype and occupational VGDF exposure in non-smokers; seven of the nine participants with high-level VGDF exposure and PiMZ genotype were ever smokers. An alternative explanation is that smoking may be a prerequisite for the interaction between PiMZ genotype and VGDF exposure to occur as it increases inflammation and therefore elastase activity in the lung.

CONCLUSION

The present findings suggest that interaction between causative factors plays an important role in the nature of lung function decline. Genetic predisposition for intermediate AAT deficiency, in combination with smoking, may increase susceptibility to occupational exposure-related decline in lung function. The interaction between PiMZ genotype and outdoor PM10 needs to be reassessed in larger studies.

Author affiliations
[1]Swiss Tropical and Public Health Institute, Basel, Switzerland
[2]University of Basel, Basel, Switzerland
[3]Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts, USA
[4]Center for Diagnosis of Inherited Alpha1-antitrypsin Deficiency, Institute for Respiratory Disease, IRCCS San Matteo Hospital Foundation, University of Pavia, Pavia, Italy
[5]Institute for Risk Assessment Sciences, Utrecht, Netherlands
[6]Occupational Medicine, Suva, Luzern, Switzerland
[7]Department of Internal Medicine, University Hospital Basel, Basel, Switzerland
[8]University Hospital of Geneva, Geneva, Switzerland
[9]University Hospital Zurich, Zurich, Switzerland

REFERENCES

1 Lange P, Parner J, Vestbo J, et al. A 15-year follow-up study of ventilatory function in adults with asthma. N Engl J Med 1998;339:1194–200.

2   Fletcher C, Peto R, Tinker C, et al. The natural history of chronic bronchitis and emphysema: An eight-year study of early chronic obstructive lung disease in working men in London. New York: Oxford University Press, 1976.

3   Balmes J, Becklake M, Blanc P, et al. American thoracic society statement: occupational contribution to the burden of airway disease. Am J Respir Crit Care Med 2003;167:787–97.

4   Eisner MD, Anthonisen N, Coultas D, et al. An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010; 182:693–718.

5   Downs SH, Schindler C, Liu LJ, et al. Reduced exposure to PM10 and attenuated age-related decline in lung function. N Engl J Med 2007;357:2338–47.

6   Schikowski T, Mills IC, Anderson HR, et al. Ambient air pollution—a cause for COPD? Eur Respir J Published Online First: 7 March 2013. doi:10.1183/09031936.00100112

7   DeMeo DL, Silverman EK. Alpha1-antitrypsin deficiency. 2: genetic aspects of alpha (1)-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. Thorax 2004;59:259–64.

8   Luisetti M, Seersholm N. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax 2004;59:164–9.

9   Ferrarotti I, Thun GA, Zorzetto M, et al. Serum levels and genotype distribution of α1-antitrypsin in the general population. Thorax 2012;67:669–74.

10 Dahl M, Hersh CP, Ly NP, et al. The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J 2005;26:67–76.

11 Hersh CP, Dahl M, Ly NP, et al. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax 2004;59:843–9.

12 Mehta AJ, Miedinger D, Keidel D, et al. Occupational exposure to dusts, gases, and fumes and incidence of chronic obstructive pulmonary disease in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults. Am J Respir Crit Care Med 2012;185:1292–300.

13 Thun GA, Ferrarotti I, Imboden M, et al. SERPINA1 PiZ and PiS heterozygotes and lung function decline in the SAPALDIA cohort. PLoS One 2012;7:e42728.

14 Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, et al. Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991–2003: methods and characterization of participants. Soz Praventivmed 2005;50:245–63.

15 Martin BW, Ackermann-Liebrich U, Leuenberger P, et al. SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on air pollution and lung diseases in adults. Soz Praventivmed 1997;42:67–84.

16 American Thoracic Society. Standardization of spirometry, 1994 update. Am J Respir Crit Care Med 1995;152:1107–36.

17 Burney PG, Luczynska C, Chinn S, et al. The European Community Respiratory Health Survey. Eur Respir J 1994;7:954–60.

18 Künzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV1 due to technician, team, device and subject in an eight centre study: three quality control studies in SAPALDIA. Swiss Study on Air Pollution and Lung Disease in Adults. Eur Respir J 1995;8:371–6.

19 Künzli N, Kuna-Dibbert B, Keidel D, et al. Longitudinal validity of spirometers—a challenge in longitudinal studies. Swiss Med Wkly 2005;135:503–8.

20 International Labour Office. International standard classification of occupations: ISCO-88. Geneva: International Labour Organization, 1990.

21 Sunyer J, Kogevinas M, Kromhout H, et al. Pulmonary ventilatory defects and occupational exposures in a population-based study in Spain: Spanish Group of the European Community Respiratory Health Survey. Am J Respir Crit Care Med 1998;157:512–7.

22 Matheson MC, Benke G, Raven J, et al. Biological dust exposure in the work-place is a risk factor for chronic obstructive pulmonary disease. Thorax 2005;60:645–51.

23 Liu LJ, Curjuric I, Keidel D, et al. Characterization of source-specific air pollution exposure for a large population-based Swiss cohort (SAPALDIA). Environ Health Perspect 2007;115:1638–45.

24 Piitulainen E, Tornling G, Eriksson S. Effect of age and occupational exposure to airway irritants on lung function in non-smoking individuals with alpha 1-antitrypsin deficiency (PiZZ). Thorax 1997;52:244–8.

25 Mayer AS, Stoller JK, Bucher Bartelson B, et al. Occupational exposure risks in individuals with PI*Z alpha(1)-antitrypsin deficiency. Am J Respir Crit Care Med 2000;162:553–8.

26 Pierre F, Pham QT, Mur JM, et al. Respiratory symptoms and pulmonary function in 871 miners according to Pi phenotype: a longitudinal study. Eur J Epidemiol 1988;4:39–44.

27 Sigsgaard T, Brandslund I, Omland O, et al. S and Z alpha1-antitrypsin alleles are risk factors for bronchial hyperresponsiveness in young farmers: an example of gene-environment interaction. Eur Respir J 2000;16:50–5.

28 Banauch GI, Brantly M, Izbicki G, et al. Accelerated spirometric decline in New York City firefighters with alpha1-antitrypsin deficiency. Chest 2010;138:1116–24.

29 Blanc PD, Iribarren C, Trupin L, et al. Occupational exposures and the risk of COPD: dusty trades revisited. Thorax 2009;64:6–12.

30 Boggia B, Farinaro E, Grieco L, et al. Burden of smoking and occupational exposure on etiology of chronic obstructive pulmonary disease in workers of Southern Italy. J Occup Environ Med 2008;50:366–70.

31 de Meer G, Kerkhof M, Kromhout H, et al. Interaction of atopy and smoking on respiratory effects of occupational dust exposure: a general population-based study. Environ Health 2004;3:6.

32 Senn O, Russi EW, Imboden M, et al. Alpha1-antitrypsin deficiency and lung disease: risk modification by occupational and environmental inhalants. Eur Respir J 2005;26:909–17.

33 de Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. Chest 2002;122:1818–29.

34 National Institute for Occupational Safety and Health (NIOSH). Genetics in the workplace—implications for occupational safety and health. DHHS Publ. No. 2010-101, US Dep. Health Hum. Serv., Public Health Serv., Cent. Dis. Control Prev., Washington, DC. 2010.

35 Schulte P, Howard J. Genetic susceptibility and the setting of occupational health standards. Annu Rev Public Health 2011;32:149–59.

36 Christiani DC, Mehta AJ, Yu CL. Genetic susceptibility to occupational exposures. Occup Environ Med 2008;65:430–6.

37 Curjuric I, Imboden M, Schindler C, et al. HMOX1 and GST variants modify attenuation of FEF25–75% decline due to PM10 reduction. Eur Respir J 2010;35:505–14.

38 Burgel PR. The role of small airways in obstructive airway diseases. Eur Respir Rev 2011;20:23–33.

39 Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J 2005;26:319–38.

ONLINE SUPPLEMENTARY MATERIAL

Interactions between SERPINA1 PiMZ Genotype, Occupational Exposure, and Lung Function Decline

Mehta AJ, Thun GA, Imboden M, Ferrarotti I, Keidel D, Künzli N, Kromhout H, Miedinger D, Phuleria H, Rochat T, Russi EW, Schindler C, Schwartz J, Vermeulen R, Luisetti M, Probst-Hensch NM and SAPALDIA-team.

D. Stolz (p), P Straehl (exp), JM Tschopp (p), A von Eckardstein (cc), JP Zellweger (p), E Zemp Stutz (e).

Scientific team at coordinating centers: M Adam (e/g), PO Bridevaux (p), D Carballo (c), E Corradi (e), I Curjuric (e), J Dratva (e), A Di Pasquale (s), L Grize (s), D Keidel (s), S Kriemler (pa), A Kumar (g), M Imboden (g), N Maire (s), A Mehta (e), F Meier (e), H Phuleria (exp), E Schaffner (s), GA Thun (g) A Ineichen (exp), M Ragettli (e), M Ritter (exp), T Schikowski (e), G Stern (pd), M Tarantino (s), M Tsai (e), M Wanner (pa).

(a) allergology, (c) cardiology, (cc) clinical chemistry, (e) epidemiology, (exp) exposure, (g) genetic and molecular biology, (m) meteorology, (n) nutrition, (o) occupational health, (p) pneumology, (pa) physical activity, (pd) pediatrics, (s) statistics

Local fieldworkers : Aarau: S Brun, G Giger, M Sperisen, M Stahel, Basel: C Bürli, C Dahler, N Oertli, I Harreh, F Karrer, G Novicic, N Wyttenbacher, Davos: A Saner, P Senn, R Winzeler, Geneva: F Bonfils, B Blicharz, C Landolt, J Rochat, Lugano: S Boccia, E Gehrig, MT Mandia, G Solari, B Viscardi, Montana: AP Bieri, C Darioly, M Maire, Payerne: F Ding, P Danieli A Vonnez, Wald: D Bodmer, E Hochstrasser, R Kunz, C Meier, J Rakic, U Schafroth, A Walder.

Administrative staff: C Gabriel, R Gutknecht.

METHODS

Study population and design

Online supplementary figure E1 outlines the inclusion criteria for this analysis. Of the 9,651 original study participants, there were 7,406 participants with complete information on spirometry, exposure, and covariates, at baseline. 6,268 of the 7,406

participants (84.6%) were also present at the follow-up survey, of which we further

excluded 1,866 individuals without completed spirometry, or information on

exposures and covariates at the follow-up survey. Additionally, we excluded 342

participants with no SERPINA1 genotyping information available, and 321

participants who were classified as having PiZZ, PiSZ, or PiMS genotype.

Spirometry testing

Identical spirometers (Sensormedics model 2200, Yorba Linda, California, USA) and

protocols were used for both surveys; comparability was assessed before and after each

survey.[E1, E2] Three to eight forced expiratory lung function manoeuvres were

performed, and at least two acceptable measurements of forced vital capacity (FVC) and

forced expiratory volume in one second (FEV_1) were obtained.

Respiratory questionnaire

Presence of asthma was defined by reporting positive to the question, "Have you ever had

asthma?" and, if yes "Was this confirmed by a doctor?" either at the baseline or follow-

up survey. Participants reporting positive to both of these questions were excluded from

the analysis.

Occupational exposure assessment

Current job titles reported by participants at baseline, and all jobs titles reported

during the follow-up period, were standardized according to the International

Standard Classification of Occupations (ISCO-88) code's four-digit

classification.[E3] The ISCO-88 classification contains 390 occupational titles and

puts more emphasis on the skills of the job than type of industry that the job belongs

to.[E3] The current job titles were then linked by ISCO-88 classification to the

ALOHA general population job-exposure matrix (GPJEM) for COPD,[E4, E5] and

were assigned scores for high, low or no exposure (2, 1, or 0, respectively) for biological dust, mineral dust, gases/fumes, and either vapours, gases, dusts, or fumes (VGDF). VGDF was a composite variable based on the highest of the assigned scores for biological dusts, mineral dusts, and gases/fumes exposures. Participants classified as having no exposure to all exposure types were classified as unexposed to VGDF. For this analysis, we only had sufficient power to evaluate occupational exposure to VGDF. Participants with either low-level of VGDF exposure or no exposure were combined as a single reference comparison group because our previous analysis observed null associations between low-level occupational exposures and risk of COPD.[E6] Cumulative exposures at baseline and during follow-up were also estimated by multiplying VGDF exposure level and the number of years worked in the specific job (unit, years), and was weighted by exposure level (1 for low exposure, 4 for high exposure) on an exponential scale to allow for the combination of high and low exposure groups.[E4]

REFERENCES

E1 Künzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV1 due to technician, team, device and subject in an eight centre study: three quality control studies in SAPALDIA. Swiss Study on Air Pollution and Lung Disease in Adults. Eur Respir J 1995;8:371-6.

E2 Künzli N, Kuna-Dibbert B, Keidel D, et al. Longitudinal validity of spirometers - a challenge in longitudinal studies. Swiss Med Wkly 2005;135:503-8.

E3 International Labour Office. International standard classification of occupations: ISCO-88. Geneva: International Labour Organization, 1990.

E4 Matheson MC, Benke G, Raven J, et al. Biological dust exposure in the work-place is a risk factor for chronic obstructive pulmonary disease. Thorax 2005;60:645-51.

E5 Sunyer J, Kogevinas M, Kromhout H, et al. Pulmonary ventilatory defects and occupational exposures in a population- based study in Spain: Spanish Group of the European Community Respiratory Health Survey. Am J Respir Crit Care Med 1998;157:512–7.

E6 Mehta AJ, Miedinger D, Keidel D, et al. Occupational exposure to dusts, gases, and fumes and incidence of chronic obstructive pulmonary disease in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults. Am J Respir Crit Care Med 2012;185:1292-300.

Figure E1 SAPALDIA study participants included in present analysis

SAP1 (1991) ⎢ **N = 9,651**

N = 7,406 **Exclusion criteria (SAP1):**
Incomplete spirometry, exposure, and covariates (n=2,245)

n = 1,138 **Loss to follow-up**

SAP2 (2001-2003) **N = 6,268** **Exclusion criteria (SAP2):**
Incomplete spirometry, exposure, and covariates (n=1,866)
No genotype (n=342)
PiSZ or PiZZ genotype (n=16)
PiMS genotype (n=305)

N = 3,739 **Participants included in present analysis**
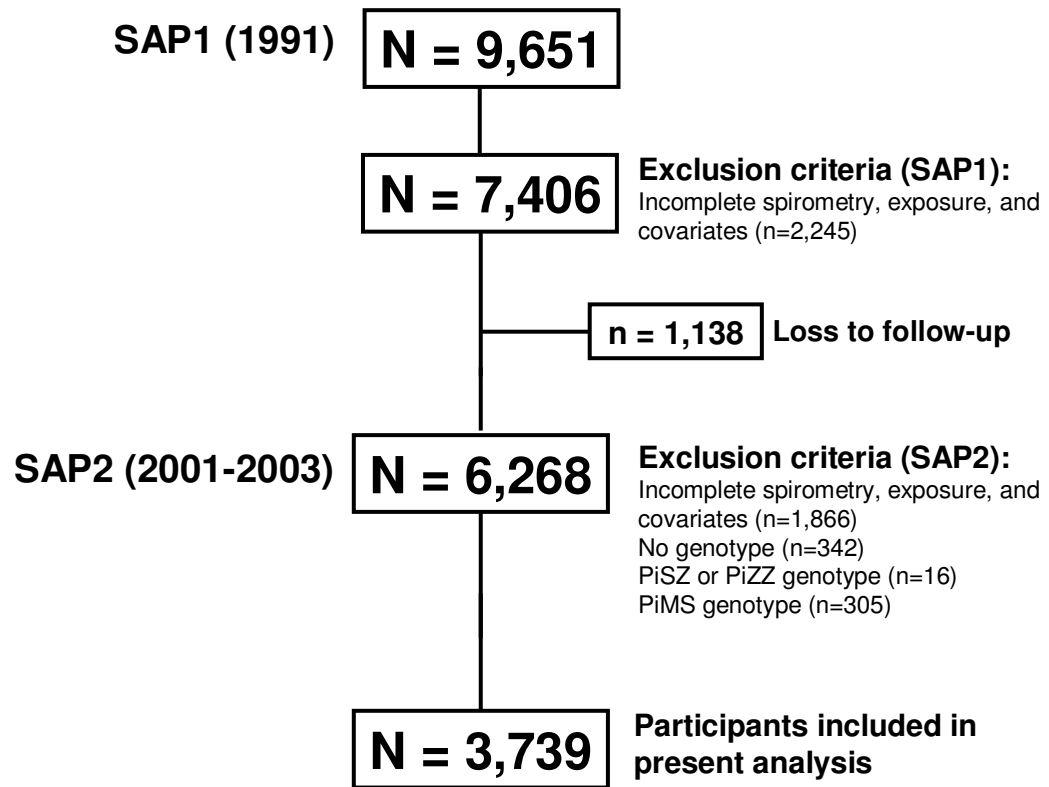
Table E1 Characteristics of participants and non-participants at baseline survey

Characteristics	Participants (n=3,739)	Non-participants (n=3,667)	P-value
FEV_1, L, mean (SD)	3.64 (0.81)	3.44 (0.86)	<0.0001
FVC, L, mean (SD)	4.61 (1.02)	4.37 (1.04)	<0.0001
FEV_1/FVC, %, mean (SD)	79.2 (7.5)	79.0 (8.4)	0.15
FEF_{25-75}, L/s, mean (SD)	3.50 (1.22)	3.33 (1.27)	<0.0001
Age, years, mean (SD)	40.4 (11.1)	41.5 (11.9)	<0.0001
BMI, kg/m^2, mean (SD)	23.7 (3.6)	24.1 (4.2)	<0.0001
Male, n (%)	1,892 (50.6)	1,672 (45.6)	<0.0001
Education level, n (%)			
High	675 (18.1)	555 (15.1)	
Medium	2,652 (70.9)	2,501 (68.2)	<0.0001
Low	412 (11.0)	611 (16.7)	
Daily ETS exposure in past 12 months, n (%)	1,364 (36.5)	1,424 (38.8)	0.04
Smoking status, n (%)			
Current	1,077 (28.8)	1,317 (35.9)	
Former	890 (23.8)	805 (22.0)	<0.0001
Never	1,772 (47.4)	1,545 (42.1)	
Cumulative pack-years smoked, mean (SD)	9.1 (15.7)	12.1 (18.7)	<0.0001
VGDF level, n (%)			
High	486 (13.0)	472 (12.9)	
Low	1,219 (32.6)	1,274 (34.7)	0.14
Unexposed	2,034 (54.4)	1,921 (52.4)	
Cumulative VGDF exposure , years, mean (SD)	13.2 (28.0)	13.7 (29.6)	0.51
1-year average PM_{10} prior to baseline survey, $\mu g/m^3$, mean (SD)	27.0 (9.8)	29.1 (9.9)	<0.0001

Table E2 Adjusted differences[*] in annual change in lung function in association with outdoor PM_{10} and occupational VGDF exposures during follow-up in PiMZ and PiMM carriers (ADJUSTING FOR BASELINE LUNG FUNCTION)

	N	ΔFEV_1 (ml) β (95%CI)[†]	ΔFVC (ml) β (95%CI)[†]	$\Delta FEV_1/FVC$ (%) β (95%CI)[†]	$\Delta FEF_{25-75\%}$ (ml/s) β (95%CI)[†]
ΔPM_{10} (per 10 μg/m³ increase in 10 years)					
PiMZ	97	-7 (-17, 2)	-4 (-19, 12)	-0.1 (-0.3, 0.0)	-21 (-45, 3)
PiMM	3,642	-2 (-6, 1)	-0 (-5, 5)	-0.0 (-0.1, 0.0)	-6 (-13, 2)
High-level VGDF exposure					
PiMZ	97	-4 (-22, 14)	11 (-13, 35)	-0.2 (-0.5, 0.1)	-51 (-87, -13)[‡]
PiMM	3,642	1 (-2, 4)	-2 (-6, 2)	0.0 (0.0, 0.1)	5 (-2, 11)

[*] Annual change in lung function was modelled in linear mixed regression with random intercept for area consisting of the following baseline covariates: lung function, age, age squared, sex, foreign status, height, body mass index, early respiratory infection, parental asthma, high-level education, cumulative VGDF exposure, and outdoor PM_{10}; additional covariates included smoking status through follow-up, cumulative pack-years through follow-up, daily environmental tobacco smoke exposure at follow-up, difference in body mass index over follow-up period, seasonality, ΔPM_{10}, high-level VGDF exposure during follow-up, and two-way interaction terms between environmental tobacco smoke exposure at follow-up and SERPINA1 genotype (PiMZ, PiMM as reference), and smoking status through follow-up and SERPINA1 genotype; we estimated the effect of ΔPM_{10} and high-level VGDF exposure during follow-up in PiMZ and PiMM carriers by adding interaction terms in the model.

[†] Negative signs indicate acceleration of decline in the lung function parameter with higher levels of exposure to the inhalants.

[‡] p=0.004 for interaction between SERPINA1 genotype and high-level VGDF exposure on $\Delta FEF_{25-75\%}$.

Table E3 Adjusted differences[*] in annual change in lung function in association with outdoor PM_{10} and occupational VGDF exposures during follow-up in PiMZ and PiMM carriers (WITH INVERSE PROBABILITY WEIGHTING)

	N	ΔFEV_1 (ml) β (95%CI)[†]	ΔFVC (ml) β (95%CI)[†]	$\Delta FEV_1/FVC$ (%) β (95%CI)[†]	$\Delta FEF_{25-75\%}$ (ml/s) β (95%CI)[†]
ΔPM_{10} (per 10 µg/m³ increase in 10 years)					
PiMZ	97	-4 (-14, 5)	1 (-15, 17)	-0.2 (-0.4, 0.0)	-28 (-56, 1)
PiMM	3,642	-3 (-6, 1)	1 (-3, 6)	-0.0 (-0.1, 0.0)	-12 (-21, -3)
High-level VGDF exposure					
PiMZ	97	-7 (-28, 13)	14 (-13, 42)	-0.3 (-0.7, 0.0)[‡]	-88 (-134, -41)[‡]
PiMM	3,642	1 (-2, 4)	-2 (-6, 2)	0.1 (0.0, 0.1)	6 (-1, 13)

[*] Annual change in lung function was modelled in linear mixed regression with random intercept for area consisting of the following baseline covariates: age, age squared, sex, foreign status, height, body mass index, early respiratory infection, parental asthma, high-level education, cumulative VGDF exposure, and outdoor PM_{10}; additional covariates included smoking status through follow-up, cumulative pack-years through follow-up, daily environmental tobacco smoke exposure at follow-up, difference in body mass index over follow-up period, seasonality, ΔPM_{10}, high-level VGDF exposure during follow-up, and two-way interaction terms between environmental tobacco smoke exposure at follow-up and SERPINA1 genotype (PiMZ, PiMM as reference), and smoking status through follow-up and SERPINA1 genotype; we estimated the effect of ΔPM_{10} and high-level VGDF exposure during follow-up in PiMZ and PiMM carriers by adding interaction terms in the model.

[†] Negative signs indicate acceleration of decline in the lung function parameter with higher levels of exposure to the inhalants.

[‡] $p<0.0001$ and $p=0.03$ for interactions between SERPINA1 genotype and high-level VGDF exposure on $\Delta FEF_{25-75\%}$ and $\Delta FEV_1/FVC$, respectively.

Table E4 Adjusted differences[*] in annual change in lung function in association with cumulative VGDF exposure during follow-up (per 10 years) in PiMZ and PiMM carriers

	N	ΔFEV_1 (ml) β (95%CI)[†]	ΔFVC (ml) β (95%CI)[†]	$\Delta FEV_1/FVC$ (%) β (95%CI)[†]	$\Delta FEF_{25-75\%}$ (ml/s) β (95%CI)[†]
PiMZ	97	-1 (-7, 4)	5 (-2, 13)	-0.1 (-0.2, -0.0)[‡]	-25 (-38, -12)[‡]
PiMM	3,642	0 (-1, 1)	-1 (-2, 1)	0.0 (-0.0, 0.0)	1 (-1, 3)

[*] Annual change in lung function was modelled in linear mixed regression with random intercept for area consisting of the following baseline covariates: age, age squared, sex, foreign status, height, body mass index, early respiratory infection, parental asthma, high-level education, cumulative VGDF exposure, and outdoor PM_{10}; additional covariates included smoking status through follow-up, cumulative pack-years through follow-up, daily environmental tobacco smoke exposure at follow-up, difference in body mass index over follow-up period, seasonality, ΔPM_{10}, cumulative VGDF exposure during follow-up (per 10 years), and two-way interaction terms between environmental tobacco smoke exposure at follow-up and SERPINA1 genotype (PiMZ, PiMM as reference), and smoking status through follow-up and SERPINA1 genotype; we estimated the effect of ΔPM_{10} and cumulative VGDF exposure during follow-up in PiMZ and PiMM carriers by adding interaction terms in the model.

[†] Negative signs indicate acceleration of decline in the lung function parameter with higher levels of exposure to the inhalants.

[‡] p=0.0001 and p=0.03 for interactions between SERPINA1 genotype and high-level VGDF exposure on $\Delta FEF_{25-75\%}$ and $\Delta FEV_1/FVC$, respectively.

6 Results: Candidate SNPs that may modify the Air Pollution Effect on Lung Function Decline

6.1 Paper 6: Follow-up on Genome-wide Main Effects: Do Polymorphisms Modify the Air Pollution Effect on Lung Function Decline in Adults?

This paper is in revision by Environment International:

Thun GA, Imboden M, Künzli N, Rochat T, Keidel D, Haun M, Schindler C, Kronenberg F, Probst-Hensch NM.

Follow-up on genome-wide main effects: do polymorphisms modify the air pollution effect on lung function decline in adults?

Gian Andri Thun, MSc[1,2], gian.thun@unibas.ch
Medea Imboden, PhD[1,2], medea.imboden@unibas.ch
Nino Künzli, MD PhD[1,2], nino.kuenzli@unibas.ch
Thierry Rochat, MD[3], thierry.rochat@hcuge.ch
Dirk Keidel, MSc[1,2], dirk.keidel@unibas.ch
Margot Haun, MSc[4], margot.haun@i-med.ac.at
Christian Schindler, PhD[1,2], christian.schindler@unibas.ch
Florian Kronenberg, MD[4], florian.kronenberg@i-med.ac.at
Nicole M Probst-Hensch, PhD[1,2], nicole.probst@unibas.ch

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland

[2] University of Basel, Basel, Switzerland

[3] Division of Pulmonary Medicine, University Hospital of Geneva, Geneva, Switzerland

[4] Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical

Pharmacology, Innsbruck Medical University, Innsbruck, Austria

Corresponding author:

Prof. Dr. Nicole Probst-Hensch
Head Unit Chronic Disease Epidemiology,
Swiss Tropical and Public Health Institute
Socinstrasse 57, P.O. Box, 4002 Basel
Switzerland

TEL   0041-61-284 83 78 (direct)
               0041-61-284 83 88 (admin)
FAX   0041-61-284 81 05

EMAIL   Nicole.Probst@unibas.ch

Abstract

Improved air quality has been found associated with attenuated age-related decline in lung function. But whether genetic polymorphisms strongly associated with lung function play a modifying role in this attenuation process has so far not been investigated.

We selected ten single nucleotide polymorphisms derived from the largest genome-wide association studies on lung function and examined whether they modified the association between the change in exposure to particulate matter $\leq$ 10 μm (ΔPM10) and lung function decline. 4310 participants from the SAPALDIA cohort provided valid spirometry measurements, a detailed pulmonary health questionnaire both at baseline and 11 years later as well as blood samples for genetic testing. Spatially and temporally resolved air pollution exposures were assigned on an individual level based on participants' residences.

Statistically significant interactions of moderate strength with ΔPM10 were detected for rs2284746. Individuals with the CC genotype profited from 21.0 ml slower annual decline of the mid expiratory flow per 10 μg/m^3 PM10 reduction over an 10-year period, while the benefits of CG and GG carriers were smaller (14.3 and 7.0 ml per year, respectively; $P_{interaction}$=0.04). The attenuated annual decline in the percentage of the forced expiratory volume in one second relative to the forced vital capacity (FEV1/FVC) was also increased with the presence of each C-allele ($P_{interaction}$=0.009). We observed

further suggestive interactions of similar magnitude in never-smokers, but none of the

results would remain statistically significant after correction for multiple testing.

We could not find strong evidence that lung function benefits from improved air quality

are modified by polymorphisms associated with lung function level in large meta-

analyzed genome-wide association studies.

Key words: air pollution, lung function, genetic polymorphism, gene-environment

interaction.

3

1. Introduction

Air pollution, a term describing a complex mixture of chemicals and particles in the ambient air, could be linked to several adverse health outcomes and chronic diseases (Chen et al., 2008). Adverse short-term effects of air pollution on respiratory health are well documented (Brunekreef et al., 1995), and evidence is also strong for a long-term effect of air pollution on slowing down growth of lung function in children. This results in deficits by the time respiratory function starts to level off and ultimately decline (Gotschi et al., 2008). Evidence for adverse long-term effects of air pollution on lung function in adults is more limited (Kunzli et al., 2010). In the Swiss Study on Air Pollution and Lung Disease in Adults (SAPALDIA), a reduction of particulate matter smaller than 10 μm in diameter (PM10) resulted in an attenuation of age-related decline of two lung function measures, namely forced expiratory volume in the first second (FEV1), as well as forced expiratory flow between 25 and 75% of the forced vital capacity (FEF25-75) (Downs et al., 2007).

Since susceptibility to air pollution appears to differ among subjects of the general population, the individual genetic background is likely to play a role in the response to inhaled pollutants. Gene-environment interaction studies have been hence carried out to find such genes which alter the association between air pollution and pulmonary disease or cardiovascular outcomes (Canova et al., 2012; Zanobetti et al., 2011). Studies on lung function have concentrated so far on candidate genes which play a role in the response to

4

free radicals (oxidative stress) (Curjuric et al., 2010; Minelli et al., 2011), in inflammatory processes (Yang et al., 2005), and in cell cycle regulation (Imboden et al., 2009), assigning effect modification to variations in the genes *NQO1*, *HMOX1*, *GSTM1*, *GPX1*, *TNF*, *LTA*, *CCND1*, and *TP53*. These single candidate genes had been selected based on biological knowledge, but a more comprehensive investigation of pathways involving several inflammatory and oxidative stress genes led to limited results at either the gene or pathway level, including a suggestive interaction of the *CRISP2* gene with the change in PM10 on lung function decline (Curjuric et al., 2012).

A large meta-analysis of 23 genome-wide association studies (GWAS) on lung function yielded 34 single nucleotide polymorphisms (SNPs) that showed new evidence of association with FEV1 or with the ratio between FEV1 and the forced vital capacity (FVC) (Soler Artigas et al., 2011). In the replication phase, the ten most strongly associated SNPs were genotyped and analyzed in nine studies including SAPALDIA. We investigated whether these polymorphisms might represent candidate genes with the potential to modify the effect of air pollution exposure on lung function decline.

2. Materials and methods

2.1 Study population and sample size

Details about the design of the SAPALDIA cohort study have been described elsewhere (Ackermann-Liebrich et al., 2005; Martin et al., 1997). Briefly, a random population

sample of adults from eight areas in Switzerland was examined in 1991 (baseline) with a detailed health questionnaire and spirometry measurements. 8047 out of 9651 were followed-up 11 years later and 6058 agreed to provide blood samples for genetic testing. For the current analysis, we excluded 784 subjects because of incomplete or invalid spirometry at either time point, 30 subjects because they had lived for less than one year at their last residential address at follow-up or could not be assigned PM10 values for other reasons, and 934 subjects due to missing covariate data. This resulted in a sample size of 4310 individuals. Due to unsuccessful genotyping (see below), the sample size was further reduced by a minimum of 22 (rs2857595) and a maximum of 41 (rs1551943) subjects (Table S1). Ethical approval was obtained from the Swiss Academy of Medical Sciences and the Regional Ethics Committees. Written informed consent was obtained from all participants before health examination and biological sample collection at both surveys.

2.2 Spirometry

Spirometry was performed without bronchodilation and according to the European Community Respiratory Health Survey protocol (Burney et al., 1994) and complied with American Thoracic Society criteria (1995). Identical spirometry protocols and devices (Sensormedics model 2200, Yorba Linda, USA) were used in 1991 and 2002 (Kunzli et al., 2005). At least two acceptable and reproducible measurements of FVC and FEV1 were obtained. In the present study, we studied annual declines of FEV1, FVC, FEV1/FVC, and FEF25–75, i.e. the differences between follow-up and baseline values of

the respective parameters divided by the individual time of follow-up in years. Thus a negative value indicates an annual loss in lung function.

2.3 Air pollution exposure and covariates

Estimates of annual average exposure to PM10 for the 12 months prior to assessments were assigned based on the participant's home address. They were derived from Gaussian dispersion models with predictions for the years 1990 and 2000, which were inter- and extrapolated based on continuous data from fixed air pollution monitoring stations (Liu et al., 2007). Individual change in PM10 exposure (ΔPM10) was calculated by subtracting estimates prior to the first health assessment from the respective estimates prior to the second assessment and scaling the difference to a 10-year time span. PM10 exposure in all study areas declined throughout the study period.

Level of education, nationality, detailed information about current and past smoking habits, parental smoking, exposure to environmental tobacco smoke (ETS), and occupational exposure to vapors, gas, dust or fumes was collected through questionnaires. Height and weight were measured, and body mass index (BMI) was calculated as weight divided by squared height. Skin-prick tests were conducted at the baseline examination and participants were classified as having atopy if they reacted positively to one or more of the eight inhalant allergens tested (Martin et al., 1997).

2.4 Genotyping

Blood for DNA analysis was provided at the follow-up examination by participants consenting to genetic analyses (Ackermann-Liebrich et al., 2005). The ten highest-

7

ranking SNPs at stage 1 of a GWAS meta-analysis on lung function were selected and

chosen for replication in SAPALDIA and other studies (Soler Artigas et al., 2011). The

genotyping of the 5646 SAPALDIA subjects with complete baseline spirometry and

smoking information was done using the iPLEX Gold MassARRAY (SEQUENOM, San

Diego, USA). Unsuccessful genotyping occurred for varying numbers of DNA samples

(rs3743563, n=33; rs12477314, n=39; rs11001819, n=37; rs7068966, n=36; rs2865531,

n=43; rs2857595, n=22; rs2284746, n=35; rs1551943, n=50; rs1529672, n=28; rs1036429, n=33),

resulting in call rates exceeding 99.1% and more for each of the ten SNPs.

2.5 Statistical analysis

SNPs were tested for Hardy-Weinberg-Equilibrium (HWE) using chi-square statistics.

We based our statistical model on the previously published analysis identifying an

attenuation of age-related lung function decline in the SAPALDIA cohort study

following the reduction in PM10 exposure (Downs et al., 2007). Mixed linear regression

models were applied to calculate genotype-specific estimates of the effect of ΔPM10 on

lung function decline with 95% confidence intervals (95% CI). Reported estimates refer

to 10 µg/m^3 declines in PM10 over 10 years of follow-up. The covariates included age,

age squared, sex, height, parental smoking status, seasonal effects (sine and cosine

function of day of examination), level of education at baseline and change in level,

nationality (Swiss or other), the presence or absence of occupational exposure to vapors,

gas, dust or fumes at both examinations, smoking status at follow-up (never vs. former

vs. current smoking), pack-years up to and since baseline, cigarettes per day in both

surveys, presence or absence of baseline atopy, PM10 at baseline, BMI at the first survey

and change in BMI as well as the interaction between the two BMI parameters.
Furthermore, we adjusted for short-term PM10 exposure considering the three days prior
to follow-up examination. Random effects for study area were introduced to control for
clustering of residuals within areas. Two-sided P-values for interaction were calculated
assuming a co-dominant genetic model.

In the sensitivity analysis restricted to never-smokers, the same covariates were
considered except for replacing the smoking variables by exposure to ETS at baseline and
during follow-up. A second sensitivity analysis replaced the annual declines of lung
function by the percent change of the annual decline.

Analyses were conducted with STATA version 12.1. P-values < 0.05 were interpreted as
statistically significant for main and interaction effects. Multiple testing was additionally
considered using Bonferroni correction, i.e. by dividing the significance level by the
number of SNPs which were investigated (n=10). The different lung function measures as
outcomes were not considered for multiple testing as they are not independent of one
another.

3. Results

Characteristics of the analyzed study population, as well as of the proportion of never-
smokers, are shown in Table 1. Never-smokers were younger on average, had a higher
percentage of females, exhibited better lung function and had a lower baseline air

pollution level assigned to. Apart from the education level, they differed in all covariates

of interest from the rest of the study population (i.e. from the ever-smokers).

Table 1 Distribution of characteristics in the study population and in never-smokers

	All	never-smokers	P[a]
Number of participants	n=4310	n=2020	
Outcomes of interest			
ΔFEV1, ml/y, mean (SD)	-35.7 (29.7)	-32.9 (28.7)	<0.001
ΔFEF25-75, ml/y, mean (SD)	-71.8 (64.4)	-68.1 (62.6)	<0.001
ΔFVC, ml/y, mean (SD)	-24.7 (40.0)	-22.2 (37.8)	0.001
$\Delta FEV_1/FVC$, $(y^{-1})*10^3$, mean (SD)	-4.1 (4.9)	-3.9 (4.8)	0.01
Covariates of interest at baseline			
Female, n (%)	2279 (52.9)	1226 (60.7)	<0.001
Age, years, mean (SD)	41.4 (11.2)	40.8 (11.9)	0.02
BMI, kg/m^2, mean (SD)	23.7 (3.6)	23.6 (3.5)	0.007
Workplace exposure to VGDF[b], n (%)	1304 (30.3)	527 (26.1)	<0.001
High education level, n (%)	1205 (28.0)	570 (28.2)	0.72
PM10, $\mu g/m^3$, mean (SD)	27.2 (9.7)	26.9 (9.6)	0.01
Covariates of interest through/at follow-up			
ΔBMI, kg/m^2, mean (SD)	2.1 (2.2)	2.2 (2.1)	0.003
Current smokers at follow-up, n (%)	944 (21.9)	0	-
Never-smokers at follow-up, n (%)	2020 (46.9)	2020 (100.0)	-

[a] P-values derive from Chi-square and Mann-Whitney tests comparing distributions of categorical and continuous variables, respectively, between never and ever-smokers.

[b] VGDF, vapors, gas, dust and fumes.

All SNPs were in HWE (P≥0.01) and showed minor allele frequencies (MAF) very close to or over 15% (Supplementary Table S1). In this study sample, a 10 μg/m^3 reduction in the home outdoor PM10 concentration over a 10-year period reduced the annual rate of decline on average by 3.7 ml (95% CI 0.3 to 7.2) for FEV1, 0.5 ml (-4.1 to 5.0) for FVC, 12.1 ml (4.3 to 19.8) for FEF25-75, and 6.2*10^{-4} (2.9*10^{-5} to 1.2*10^{-3}) for FEV1/FVC. Genotype-specific estimates of the effects of a 10 μg/m^3 change in PM10 on lung function decline are given in Table 2. The statistically significant effects of air pollution change on lung function decline, in particular concerning mid-term flow and to a lesser extent FEV1 (Downs et al., 2007), are for most SNPs visible in the genotype strata with the largest number of individuals. Genotype-specific differences in effect estimates were only moderately significant for one of the polymorphisms. While CC carriers of rs2284746 could profit of 21.0 ml slower annual loss of FEF25-75 per 10 μg/m^3 PM10 reduction, carriers of CG and GG genotypes had a smaller benefit (attenuation of 14.3 and 7.0 ml per year, respectively). A similar trend could be observed in terms of the change in FEV1/FVC decline, whereby carriers of two G alleles did not show any profit at all. P-values for interaction reached statistical significance only if correction for multiple testing was not taken into account (P=0.04 for ΔFEF25-75 and P=0.009 for ΔFEV1/FVC). When replacing absolute lung function changes with relative ones (%-change in lung function), results were similar, confirming the observed interaction and not pointing to any additional ones.

Table 2. Genotype-specific effect estimates[a] for annual change in lung function decline by improved PM10 exposure.

	N	ΔFEV1 (ml)	P	ΔFVC (ml)	P	ΔFEF25-75 (ml/s)	P	ΔFEV₁/FVC (*10⁴)	P
rs3743563, GG[b]	2995	-4.3	0.03	-1.2	0.63	-14.3	0.001	-6.3	0.06
rs3743563, GA	1159	-5.2	0.06	0.1	0.97	-15.6	0.01	-9.8	0.04
rs3743563, AA	124	5.2	0.48	14.9	0.13	-4.7	0.78	-13.4	0.30
rs12477314, CC	2650	-5.1	0.01	-0.9	0.73	-14.8	0.001	-8.7	0.01
rs12477314, CT	1454	-3.0	0.20	-0.0	1.00	-13.7	0.01	-4.7	0.25
rs12477314, TT	167	-1.6	0.78	0.8	0.91	-12.6	0.33	-7.4	0.45
rs11001819, GG	1273	-6.2	0.01	-1.6	0.62	-20.7	<0.001	-8.5	0.04
rs11001819, GA	2092	-3.1	0.15	-0.4	0.89	-10.4	0.03	-5.2	0.16
rs11001819, AA	909	-5.1	0.07	-0.5	0.88	-15.3	0.02	-10.5	0.03
rs7068966, TT	1275	-4.5	0.08	0.2	0.96	-16.0	0.007	-10.2	0.02
rs7068966, TC	2046	-5.3	0.01	-2.3	0.42	-12.9	0.007	-5.8	0.11
rs7068966, CC	956	-1.7	0.53	2.4	0.50	-16.4	0.008	-7.2	0.12
rs2865531, AA	1450	-5.1	0.03	0.5	0.88	-15.0	0.004	-11.2	0.005
rs2865531, AT	2050	-4.2	0.04	-1.6	0.56	-13.7	0.005	-5.0	0.17
rs2865531, TT	773	-1.9	0.53	-0.5	0.91	-12.6	0.07	-3.5	0.51
rs2857595, GG	3048	-3.6	0.07	0.4	0.88	-14.1	0.002	-7.9	0.02
rs2857595, GA	1126	-6.1	0.02	-3.3	0.33	-15.0	0.01	-5.3	0.24
rs2857595, AA	114	0.7	0.92	4.1	0.65	-0.1	1.00	-7.1	0.55
rs2284746, CC	1152	-4.9	0.05	2.0	0.55	**-21.0[c]**	**<0.001**	**-14.1**	**0.001**
rs2284746, CG	2153	-4.5	0.03	-1.7	0.54	**-14.3**	**0.003**	**-6.7**	**0.07**
rs2284746, GG	973	-3.5	0.20	-1.6	0.66	**-7.0**	**0.26**	**-0.5**	**0.91**
rs1551943, GG	2543	-4.8	0.02	-0.5	0.86	-16.7	<0.001	-9.1	0.009
rs1551943, GA	1512	-4.2	0.07	-1.4	0.66	-12.0	0.03	-4.4	0.28
rs1551943, AA	214	0.1	0.98	3.0	0.68	-10.8	0.39	-8.6	0.36
rs1529672, CC	3121	-5.0	0.009	-2.1	0.41	-14.6	0.001	-6.3	0.06
rs1529672, CA	1060	-0.8	0.77	5.5	0.12	-11.9	0.05	-10.5	0.02

12

	N								
rs1529672, AA	99	-10.3	0.19	-9.8	0.35	-19.1	0.29	-2.1	0.88
rs1036429, CC	2595	-4.6	0.02	-0.5	0.85	-14.2	0.002	-8.2	0.02
rs1036429, CT	1465	-3.8	0.12	-0.6	0.86	-16.4	0.003	-5.8	0.16
rs1036429, TT	219	-5.6	0.24	-2.6	0.68	-9.9	0.36	-7.4	0.37

[a] Negative effect estimates describe an attenuated annual decline in lung function per 10 µg/m^3 PM10 reduction in 10 years.

[b] Genotypes refer to the + strand and are tabled according to frequency.

[c] Interaction between genotype and change in PM10 is shown in bold if statistically significant without correction for multiple testing (0.005<P<0.05). None of the interactions between genotype and change in PM10 would reach statistical significance after correction for multiple testing (P<0.005).

Table 3. Genotype-specific effect estimates[a] for annual change in lung function decline by improved PM10 exposure in never-smokers.

	N	ΔFEV1 (ml)	P	ΔFVC (ml)	P	ΔFEF25-75 (ml/s)	P	ΔFEV$_1$/FVC (*10^4)	P
rs3743563, GG[b]	1398	-5.2	0.05	-3.3	0.34	-16.0	0.009	-6.3	0.18
rs3743563, GA	555	-5.1	0.20	-1.2	0.81	-16.0	0.08	-8.6	0.21
rs3743563, AA	51	-1.0	0.93	9.3	0.52	-23.0	0.37	-22.5	0.25
rs12477314, CC	1258	-6.7	0.02	-3.6	0.31	-17.8	0.005	-7.7	0.11
rs12477314, CT	681	-2.8	0.42	-1.7	0.71	-12.6	0.11	-5.8	0.33
rs12477314, TT	62	1.0	0.90	4.7	0.65	-17.2	0.35	-8.1	0.56
rs11001819, GG	601	**-10.3**[c]	**0.002**	-6.3	0.14	-25.5	0.001	-8.2	0.16
rs11001819, GA	978	**-3.8**	**0.20**	-1.9	0.62	-11.2	0.10	-6.3	0.22
rs11001819, AA	418	**0.2**	**0.97**	2.0	0.71	-15.6	0.10	-8.5	0.24
rs7068966, TT	595	**-9.6**	**0.007**	-2.0	0.66	**-31.5**	**<0.001**	**-21.1**	**0.001**
rs7068966, TC	962	**-4.5**	**0.13**	-4.6	0.23	**-8.9**	**0.19**	**-0.5**	**0.92**
rs7068966, CC	444	**0.1**	**0.99**	0.9	0.85	**-10.8**	**0.23**	**-2.9**	**0.67**
rs2865531, AA	678	-3.8	0.25	2.0	0.63	-13.3	0.07	-12.1	0.03
rs2865531, AT	962	-6.7	0.03	-7.1	0.07	-16.5	0.02	-2.3	0.67
rs2865531, TT	365	-4.5	0.28	-1.4	0.79	-20.3	0.03	-9.0	0.21
rs2857595, GG	1401	**-2.7**	**0.34**	-0.8	0.81	*-11.7*[d]	*0.07*	-6.0	0.22
rs2857595, GA	565	**-11.1**	**0.002**	-8.1	0.07	*-24.2*	*0.002*	-8.7	0.15
rs2857595, AA	43	**-3.7**	**0.71**	4.7	0.71	*-31.7*	*0.16*	-12.5	0.47
rs2284746, CC	543	-6.1	0.09	0.2	0.97	-23.3	0.004	**-17.3**	**0.005**
rs2284746, CG	1015	-6.3	0.04	-5.6	0.15	-16.1	0.02	**-4.8**	**0.36**
rs2284746, GG	451	-2.9	0.43	-2.0	0.68	-8.4	0.32	**0.8**	**0.91**
rs1551943, GG	1171	-3.7	0.20	**1.2**	**0.75**	-17.9	0.006	*-12.2*	*0.01*
rs1551943, GA	731	-7.4	0.03	**-8.7**	**0.04**	-12.4	0.10	*2.2*	*0.70*
rs1551943, AA	100	-8.6	0.28	**-7.1**	**0.49**	-23.4	0.20	*-11.9*	*0.39*

14

rs1529672, CC	1444	-5.8	0.03	-3.3	0.35	-17.9	0.004	-7.6	0.11
rs1529672, CA	520	-2.6	0.49	-0.1	0.98	-8.8	0.31	-5.7	0.39
rs1529672, AA	40	-12.4	0.32	-21.0	0.19	-21.0	0.46	8.5	0.69
rs1036429, CC	1227	-5.6	0.04	-2.2	0.53	-16.6	0.008	-9.3	0.05
rs1036429, CT	669	-3.2	0.36	-2.9	0.52	-13.9	0.08	-1.2	0.84
rs1036429, TT	107	-18.9	0.04	-11.5	0.32	-37.3	0.07	-25.1	0.11

[a] Negative effect estimates describe an attenuated annual decline in lung function per 10 $\mu g/m^3$ PM10 reduction in 10 years.

[b] Genotypes refer to the + strand and are tabled according to frequency.

[c] Interaction between genotype and change in PM10 is shown in bold if statistically significant without correction for multiple testing ($0.005 < P < 0.05$). None of the interactions between genotype and change in PM10 would reach statistical significance after correction for multiple testing ($P < 0.005$).

[d] Interaction between genotype and change in PM10 is shown in italics if borderline statistically significant without correction for multiple testing ($0.05 < P < 0.10$).

Results of equivalent analyses in never-smokers are presented in Table 3. While the strongest

interaction with decline in air pollution exposure was again found for rs2284746 (P=0.01 on

ΔFEV1/FVC), several other interactions of moderate strength could be observed. For

rs7068966, we found suggestive interactions with air pollution change on three of the four

spirometric parameters considered. But like for the whole sample, none remained significant

after correction for multiple testing. Using %-changes of lung function as outcome measures

pointed to the same moderate modifications of the effect of air pollution decline, apart from

those by rs2857595, which were absent.

4. Discussion

The effect of reduced PM10 in the ambient air on longitudinal lung function was not

substantially modified by SNPs selected for their strong association with lung function. In the

whole study sample, we only found suggestive evidence for an interaction with rs2284746, an

intronic SNP in the microfibrillar-associated protein 2 gene (*MFAP2*), which previously

showed consistent gene main effects on FEV1/FVC (Soler Artigas et al., 2011).

The G allele was associated with lower lung function level in the previous GWAS meta-

analysis than the C allele (Soler Artigas et al., 2011), and in our gene-environment interaction

analysis, we identified subjects carrying the G allele to be least likely to benefit from

improved air quality. The genotype-specific effect estimates were of similar magnitude in the

subsample of never-smokers, which does preclude smoking-related interpretations of our

findings. The inclusion of baseline adjustment for lung function did also not materially alter

estimates of effect modification, which argues against a confounding of our findings by

16

differences in baseline lung function between genotypes. The rs2284746 variant has further

been shown to be associated with height (Lango Allen et al., 2010), but its inverse association

with lung function made it unlikely to merely be a consequence of the association with height

(Soler Artigas et al., 2011). *MFAP2*, encoding for a major antigen of elastin-associated

microfibrils (Gibson et al., 1986), seems to be important in the extracellular matrix function,

and inactivation led to complex phenotypes in several organ systems in mice, including

altered body size, fat deposition or wound healing (Weinbaum et al., 2008). The gene showed

expression in several tissues of the human respiratory system, but no transcripts were found

for example in human peripheral blood mononuclear cells (Soler Artigas et al., 2011).

Nevertheless, the observed interaction was modest and did not remain significant after

adjustment for multiple testing. In a previous candidate gene-environment interaction analysis

in the SAPALDIA cohort study with similar inclusion criteria and sample size,

polymorphisms in cell fate controlling genes showed very comparable effect estimates, but p-

values for interaction were slightly stronger as they were optimized for the best-fitted

underlying genetic model (Imboden et al., 2009).

The SNPs assessed in this work are common variants most consistently associated with lung

function in a meta-analysis including 48,201 participants of European ancestry across 23

GWAS. None of these SNPs interacted with smoking (Soler Artigas et al., 2011), and an

attempt to find additional variants by performing a joint test for SNP main effects and

interaction effects with smoking in the same study sample did not reveal promising genetic

loci with statistically significant interactions with smoking (Hancock et al., 2012). However,

interactions with environmental or lifestyle factors may explain part of the heritability of the

different lung function measures, which are estimated at around 50% (Wilk et al., 2000). This

number is much higher than the maximal 10% which can be explained by combining the

17

separate SNP main effect estimates having been reported up to now. In the present study, air pollution showed a tendency to modify effects of some of the investigated SNPs especially in never-smokers, a finding that supports the assumption that different mechanisms and gene pathways mediate air pollution and smoking effects (Curjuric et al., 2012). Of concern is, however, that SNPs being associated with cross-sectional lung function do not appear to be associated with the decline of lung function as measured in longitudinal studies (Imboden et al., 2012). The SNPs described here did not show association with lung function change in a subset of SAPALDIA for which we had genome-wide data available. GWAS on lung function decline, however, have so far not reported any significant results that could be confirmed in independent replication studies, probably in part due to insufficient statistical power or more complex interactions of genetic and environmental factors as main determinants of loss of lung function (Hansel et al., 2013; Imboden et al., 2012).

Our study benefits from a large population-based sample with high-quality data on health parameters, lifestyle factors and genotypes, as well as validated data on exposure to air pollution (PM10) on an individual scale. As shown repeatedly in former SAPALDIA publications (Curjuric et al., 2010), follow-up participants were healthier than other recruited individuals at baseline. This could be of concern if the genetic profile influenced the participation rate. Population stratification is most likely not a relevant factor, since length of residency prior to the first survey and language constraints led to a predominantly Swiss sample, non-Swiss citizenship was adjusted for, and genotype frequencies according to language regions did not show any differences. Models were adjusted for a range of possible confounders that were derived in a careful attempt to model the impact of air pollution change on lung function change (Downs et al., 2007). Lung function change was determined from only two spirometric measurements, which is a limitation. As the same spirometers and stringent quality criteria were applied, measurement error is unlikely to play an important role

18

in our measurements, and if so, it may be randomly distributed across genotypes and thereby reduce the statistical power.

5. Conclusion

This is the first study which tested gene-air pollution interactions for polymorphisms selected on the basis of their consistent contribution to lung function and not on the basis of functional knowledge of an underlying gene. In conclusion, we found no compelling evidence for the ten agnostically identified GWAS SNPs to play a role in conferring susceptibility to air pollution exposure. In the future, functional studies must address how these SNPs act on lung function. Missing interactions with smoking and modest interactions with air pollution suggest that these SNPs point to biological mechanisms of relevance to lung function that are not or only weakly related to inhaled irritants. The genes strongly interacting with air pollution need yet to be identified, as the present study could not substantially add to the present knowledge. Gene-environment interaction studies in countries with rising or very high air pollution levels as well as agnostic procedures like genome-wide interaction studies (GWIS) should be considered in the future. This may help to identify new genes and elucidate biological mechanisms underlying susceptibility to air pollution.

Acknowledgements

References

(1995). Standardization of Spirometry, 1994 Update. American Thoracic Society. Am J Respir Crit Care Med *152*, 1107-1136.

Ackermann-Liebrich, U., Kuna-Dibbert, B., Probst-Hensch, N.M., Schindler, C., Felber Dietrich, D., Stutz, E.Z., Bayer-Oglesby, L., Baum, F., Brandli, O., Brutsche, M.*, et al.* (2005). Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed *50*, 245-263.

Brunekreef, B., Dockery, D.W., and Krzyzanowski, M. (1995). Epidemiologic studies on short-term effects of low levels of major ambient air pollution components. Environ Health Perspect *103 Suppl 2*, 3-13.

Burney, P.G., Luczynska, C., Chinn, S., and Jarvis, D. (1994). The European Community Respiratory Health Survey. Eur Respir J *7*, 954-960.

Canova, C., Dunster, C., Kelly, F.J., Minelli, C., Shah, P.L., Caneja, C., Tumilty, M.K., and Burney, P. (2012). PM10-induced hospital admissions for asthma and chronic obstructive pulmonary disease: the modifying effect of individual characteristics. Epidemiology *23*, 607-615.

Chen, H., Goldberg, M.S., and Villeneuve, P.J. (2008). A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. Reviews on environmental health *23*, 243-297.

Curjuric, I., Imboden, M., Nadif, R., Kumar, A., Schindler, C., Haun, M., Kronenberg, F., Kunzli, N., Phuleria, H., Postma, D.S.*, et al.* (2012). Different genes interact with particulate matter and tobacco smoke exposure in affecting lung function decline in the general population. PloS one *7*, e40175.

Curjuric, I., Imboden, M., Schindler, C., Downs, S.H., Hersberger, M., Liu, S.L., Matyas, G., Russi, E.W., Schwartz, J., Thun, G.A.*, et al.* (2010). HMOX1 and GST variants modify attenuation of FEF25-75% decline due to PM10 reduction. Eur Respir J *35*, 505-514.

Downs, S.H., Schindler, C., Liu, L.J., Keidel, D., Bayer-Oglesby, L., Brutsche, M.H., Gerbase, M.W., Keller, R., Kunzli, N., Leuenberger, P.*, et al.* (2007). Reduced exposure to PM10 and attenuated age-related decline in lung function. The New England journal of medicine *357*, 2338-2347.

Gibson, M.A., Hughes, J.L., Fanning, J.C., and Cleary, E.G. (1986). The major antigen of elastin-associated microfibrils is a 31-kDa glycoprotein. J Biol Chem *261*, 11429-11436.

Gotschi, T., Heinrich, J., Sunyer, J., and Kunzli, N. (2008). Long-term effects of ambient air pollution on lung function: a review. Epidemiology *19*, 690-701.

Hancock, D.B., Artigas, M.S., Gharib, S.A., Henry, A., Manichaikul, A., Ramasamy, A., Loth, D.W., Imboden, M., Koch, B., McArdle, W.L.*, et al.* (2012). Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. PLoS genetics *8*, e1003098.

21

Hansel, N.N., Ruczinski, I., Rafaels, N., Sin, D.D., Daley, D., Malinina, A., Huang, L., Sandford, A., Murray, T., Kim, Y.*, et al.* (2013). Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Human genetics *132*, 79-90.

Imboden, M., Bouzigon, E., Curjuric, I., Ramasamy, A., Kumar, A., Hancock, D.B., Wilk, J.B., Vonk, J.M., Thun, G.A., Siroux, V.*, et al.* (2012). Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol *129*, 1218-1228.

Imboden, M., Schwartz, J., Schindler, C., Curjuric, I., Berger, W., Liu, S.L., Russi, E.W., Ackermann-Liebrich, U., Rochat, T., Probst-Hensch, N.M.*, et al.* (2009). Decreased PM10 exposure attenuates age-related lung function decline: genetic variants in p53, p21, and CCND1 modify this effect. Environ Health Perspect *117*, 1420-1427.

Kunzli, N., Kuna-Dibbert, B., Keidel, D., Keller, R., Brandli, O., Schindler, C., Schweinzer, K.M., Leuenberger, P., Ackermann-Liebrich, U., and team, S. (2005). Longitudinal validity of spirometers--a challenge in longitudinal studies. Swiss Med Wkly *135*, 503-508.

Kunzli, N., Perez, L., and Rapp, R. (2010). Air quality and health. In: European Respiratory Society, Switzerland.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S.*, et al.* (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832-838.

Liu, L.J., Curjuric, I., Keidel, D., Heldstab, J., Kunzli, N., Bayer-Oglesby, L., Ackermann-Liebrich, U., Schindler, C., and team, S. (2007). Characterization of source-specific air pollution exposure for a large population-based Swiss cohort (SAPALDIA). Environ Health Perspect *115*, 1638-1645.

Martin, B.W., Ackermann-Liebrich, U., Leuenberger, P., Kunzli, N., Stutz, E.Z., Keller, R., Zellweger, J.P., Wuthrich, B., Monn, C., Blaser, K.*, 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 Praventivmed *42*, 67-84.

Minelli, C., Wei, I., Sagoo, G., Jarvis, D., Shaheen, S., and Burney, P. (2011). Interactive effects of antioxidant genes and air pollution on respiratory function and airway disease: a HuGE review. American journal of epidemiology *173*, 603-620.

Soler Artigas, M., Loth, D.W., Wain, L.V., Gharib, S.A., Obeidat, M., Tang, W., Zhai, G., Zhao, J.H., Smith, A.V., Huffman, J.E.*, et al.* (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet *43*, 1082-1090.

Weinbaum, J.S., Broekelmann, T.J., Pierce, R.A., Werneck, C.C., Segade, F., Craft, C.S., Knutsen, R.H., and Mecham, R.P. (2008). Deficiency in microfibril-associated glycoprotein-1 leads to complex phenotypes in multiple organ systems. J Biol Chem *283*, 25533-25543.

Wilk, J.B., Djousse, L., Arnett, D.K., Rich, S.S., Province, M.A., Hunt, S.C., Crapo, R.O., Higgins, M., and Myers, R.H. (2000). Evidence for major genes influencing pulmonary function in the NHLBI family heart study. Genetic epidemiology *19*, 81-94.

Yang, I.A., Holz, O., Jorres, R.A., Magnussen, H., Barton, S.J., Rodriguez, S., Cakebread, J.A., Holloway, J.W., and Holgate, S.T. (2005). Association of tumor necrosis factor-alpha polymorphisms and ozone-induced change in lung function. Am J Respir Crit Care Med *171*, 171-176.

Zanobetti, A., Baccarelli, A., and Schwartz, J. (2011). Gene-air pollution interaction and cardiovascular disease: a review. Progress in cardiovascular diseases *53*, 344-352.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Supplementary Table S1. Characteristics of the genotyped single nucleotide polymorphisms

Single Nucleotide Polymorphisms	Call rate[a] (%)	MAF[a] (%)	P(HWE)[a]	N
rs3743563, G/A[b]	99.4	16.4	0.62	4278
rs12477314, C/T	99.3	21.1	0.11	4271
rs11001819, G/A	99.3	45.4	0.12	4274
rs7068966, T/C	99.4	46.1	0.01	4277
rs2865531, A/T	99.2	42.0	0.25	4273
rs2857595, G/A	99.6	16.0	0.43	4288
rs2284746, C/G	99.4	47.9	0.56	4278
rs1551943 G/A	99.1	22.9	0.41	4269
rs1529672 C/A	99.5	14.8	0.10	4280
rs1036429 C/T	99.4	22.7	0.88	4279

[a] Call rate, minor allele frequency (MAF) and P-values for Hardy-Weinberg-Equilibrium (P(HWE)) refer to the total genotyped sample size of N=5646.

[b] Given are the major and the minor allele on the +-strand.

7 Discussion

7.1 Main Findings in a General Context

Traditionally, human geneticists explored their research domain by concentrating on rare diseases with high penetrance, making use of case-reports, studies in families, pedigrees or isolated populations. However, the major global health burden is composed of complex diseases with low penetrance, including most NCDs as well as some communicable diseases with a chronic course. Genetic aspects of such diseases have usually been experimentally studied in cell cultures or animal models by researchers trained in molecular medicine. Epidemiologists, on the other hand, have been investigating a wide range of exogenous risk factors and their associations with complex diseases and disease-relevant traits using population-based surveys, case-control or cohort studies. Genetics has usually not played an important role in their research. The present PhD work has taken place in the relatively young and thriving discipline of genetic epidemiology where human genetics meets epidemiology. Its ultimate goal is the establishment of causal links between genomic variability, environmental influences and phenotypic variability including health. By clarifying mechanisms and disease susceptibilities, this research field further helps to improve preventive action against modifiable risk factors and to identify biomarkers for risk screening.

Most projects of this PhD were related to AAT serum level, a quantitative metabolic trait influencing pulmonary health. Genetic causes for differing serum concentrations were determined and refined, the association with pulmonary health outcomes in dependence of the environment was investigated, and diagnostic values were specified. Side projects examined genetic interactions with obesity in relation to asthma and with air pollution in relation to lung function decline.

7.1.1 The Genetics of AAT

The GWAS on AAT serum level confirmed that genetic variation in *SERPINA1* is by far the most influential determinant of AAT protein concentration in the blood (chapter 5.2). Moreover, associations with common variants in the *SERPINA* locus could all be attributed to stronger associations with the well-established non-synonymous polymorphisms PiS and

PiZ, which show low frequency in the general population. Replication in an independent study sample, targeted fine mapping of the *SERPINA1* region including conditional analyses, additional SNP imputation in selected regions using most up-to-date 1000 genomes reference panels, look-ups in eQTL data bases [86] and bioinformatic prediction tools providing functional SNP annotation all contributed to strengthen and refine the original GWAS findings. The strong synthetic associations which we found in this locus are explainable by the low genetic recombination rate in this region (i.e. by the high LD in terms of D'). Interestingly, this is not a novel discovery. The order of the genes in the *SERPINA* cluster was clarified back in 1994 by assessing genetic recombination events through linkage analysis in AAT deficient families. High allelic association was already observed at that time, and the PiZ allele was found in a unique haplotype stretching over 60 kb in 97% of all PiZ carriers [185]. This is in agreement with the assumption that the PiZ allele had only one origin and might have arisen about 2000 years ago [186].

7.1.2   Causal Variants of Complex Traits, Common or Rare?

In terms of the open debate about whether the missing heritability of human traits is hidden in common variants with small effects or in rare variants with larger effects [187], the synthetic associations observed in our GWAS are a strong supporter of the potential impact of rare variants. We also conclude from our results that common variants in the *SERPINA* gene cluster which were formerly associated with lung function and emphysema [112,130] are most likely not causal. Such dependency on rare variant effects may accordingly apply to many associations of common variants found in GWAS.

Rare variants have long been acknowledged for their influence in monogenic disorders (e.g. [188]), but due to the different methodology, examples of a causal role in polygenic complex disorders have much later been reported [189,190]. From an evolutionary perspective and reflections about the distribution of allele frequencies in the genome, rare variants are more likely than common variants to be deleterious and held to low frequency by purifying selection [191,192]. Models that explain association results from current GWAS as mainly synthetic are supportive of this argument [193]. Furthermore, associated common variants were often observed in regions without functional annotation. This argument is however losing ground with the rapid progress in allocating functionality to the

genome. In fact, GWAS results were shown to be enriched in functional genetic regions, especially in 5' untranslated regions [194].

Supporters of the common disease-common variant hypothesis argue that balancing selection (and different environmental pressures in human history) may explain how a disease-associated variant became common. For instance, it was hypothesized that the PiZ allele got relatively widespread by conferring a potential advantage in the pre-antibiotic era (e.g. by the generation of polymers at sites of inflammation, which potentially attracted neutrophils leading to a stronger inflammatory response) [195]. Since smoking was less widespread and life expectancy much shorter than nowadays, the drawbacks of being PiZZ carrier might have been negligible. Many complex diseases are in fact late-onset and affect the fitness after the reproductive phase. We could therefore argue that selection against disease-associated variants becoming common is more relevant in communicable diseases, which affect the fitness earlier in life. Others being in favour of an important role for common variants mention that if effects of common variants were synthetic, the true effects of underlying rare variants would have to be strong enough to be detectable by genetic linkage analysis [196]. Moreover, for some traits like the human height, hundreds of independent common variants were found. If they all relied on stronger rare variants, the total variation explained would by far exceed the heritability [197]. Recent studies claim that they can explain a large amount of the heritability of many complex diseases by estimating the influence of thousands of common variants with very small effects [51]. The strong correlation between GWAS effect sizes of Europeans and East Asians for several diseases points also to the causality of common variants [198], but such correlation is also controversially discussed [199].

Empirical examples from targeted fine-mapping studies do not clearly favour one or the other explanation (e.g. [200-203]). Deep sequencing studies of collections of genes or whole exomes in thousands of individuals were recently facilitated by the drop in sequencing costs. They showed that almost 90% of SNPs per individual are common, but population-wise, there is an excess of rare variants compared to frequency spectrum estimates [204,205]. They arose to a large part in the past 5000 years and are geographically localized reflecting the recent explosive population growth combined with purifying selection [206]. Among those genetic variants assigned to disease-associated genes or with computationally predicted adverse function, the proportion of rare variants is elevated. These results infer

that rare variants contribute importantly to disease susceptibility. Large-scale WGS may eventually resolve this open debate, but many complex diseases and medically-relevant traits will most likely depend on both common and independently associated low-frequent or rare variants. They may show a heritability pattern like human height, i.e. common SNPs in total account for far more of the genetic variation than the statistically highly significant SNPs detected in GWAS, but cannot explain all of the genetic variation estimated by pedigree methods [207,208]. The heritability is therefore to a certain degree hidden and only to a probably smaller part missing in current GWAS results. This has recently been confirmed for COPD. In a smoking population of Non-Hispanic White and African Americans, heritability utilizing all SNPs within a GWAS was estimated at 38% [209], which is much more than what known disease-associated variants can explain ($< 10\%$), but less than what pedigree-based methods suggest (see 1.2).

7.1.3 AAT Deficiency Genotypes and Lung Function

The transfer of genetic association patterns with AAT serum concentration to according associations with lung function proved challenging (chapter 5.2). Neither the PiS nor the PiZ variant were associated with lung function in ever-smokers in meta-analyses with over 20 000 persons. This is in agreement with meta-analysed results for the general population [151,152] and indicates that the status of ever-smoking is not sufficient to put heterozygous PiS or PiZ carriers at higher risk for adverse pulmonary health than their PiMM counterparts. But the fact that we found dependency of lung function on the PiZ deficiency allele merely in patient studies strengthened again the notion that further predisposing factors interacting with this deficiency allele must exist. The confirmation that the PiMZ genotype is generally not linked to lower lung function also confers low probability to the assumption of a causal association between moderately reduced AAT levels and low lung function. Such a Mendelian randomisation approach would prove helpful since the direct measurement of an association between AAT blood levels and lung function was hampered by the fact that individuals with lung function impairment often exhibited elevated AAT levels due to its role as an inflammatory marker [136].

In the search for the additional factors putting PiMZ individuals at risk, we concentrated on environmental factors which were known to independently impact lung function. Genetic factors were not assessed since the number of possible interactions with *SERPINA1* would

be high, and neither the GWAS discovery sample nor the look-up in the eQTL database pointed to convincing candidate loci. Concentrating on inflammatory stimuli, the PiMZ genotype was associated with an accelerated $FEF_{25-75\%}$ decline in smoking and obesity subgroups from the general population (chapter 5.3). The interaction was especially strong with pack-years, showing a dose-dependent manner. Weaker, but consistent associations were also observed with the FEV_1/FVC decline. Results for participants exhibiting high CRP values (although mostly not in the clinically relevant range) strengthened the notion that PiMZ carriers might be more susceptible to inflammatory triggers with respect to lung function parameters used as indices for assessing mild airway obstruction.

In a similar approach, PiMZ genotype modified the association between lung function decline and exposure to vapours, gas, dusts and fumes, an interaction which was only observed in smokers (chapter 5.4). Interestingly, the same lung function measures were concerned, namely the $FEF_{25-75\%}$ decline and to a lesser extent the FEV_1/FVC decline. The mid expiratory flow is not a widely assessed spirometric measure (see 1.3.1), but since we have repeatedly found main and interacting effects of air pollution strongest for this mid flow parameter [172,179,210], measurement error seems an unlikely explanation. We did not observe statistically significant interaction between PiMZ genotype and air pollution change. This could be explained by the comparably modest levels of air pollution in most regions of Switzerland. Heavy air pollution like New York City firefighters experienced after September 11 attacks resulted in a FEV_1 decline dependent on AAT serum levels in a dose-dependent manner, while no such difference was observed before September 11 [178].

Taking together, these findings suggest that interaction between causative factors is an important process in the nature of lung function decline. This is in good agreement with the threshold concept related to AAT serum level and lung function. While a certain amount of inflammatory stress and therefore elevated elastase level seems to be tolerated, additional exposures and/or lack of antiprotease activity exceed this ability. In both projects, sensitivity analyses which either excluded self-reported asthmatics or which gave more weight to underrepresented groups due to a possible selection of healthy individuals (see 3.1.2) did not alter the results. Statistical power was an issue. Persistent smoking, obesity or high occupational exposure was rare in combination with PiMZ genotype. Moreover, the clinical meaning of the enlarged decline of PiMZ subjects remained unclear since the statistical power did not allow us to properly apply a definition of incident airway obstruction.

7.1.4 Reference Values for AAT Deficiency Alleles

As we have just seen, subgroups of persons with PiMZ genotype may be at increased risk for adverse lung health if certain co-factors are present. Since the quantification of AAT blood level is usually the first step towards a detection of deficiency alleles, the specification of reference values for the most common AAT deficiency genotypes is important. We performed such analyses for the first time in a population-based sample and observed narrower ranges than present in the literature (chapter 5.1). Moreover, the overlap between PiMM and PiMZ individuals was substantially smaller than previously reported. By making use of these values, we recalculated upper (92 mg/dl) and lower (49 mg/dl) limits of the concentration range corresponding to an intermediate deficiency. The latter represents thereby the protective threshold below which a decision in favour of a substitution therapy may be considered. The optimal threshold in terms of sensitivity and specificity of having at least one PiZ allele was found at 100 mg/dl. This cut-off, above which further *SERPINA1* genotyping or phenotyping seems not meaningful, at least in the absence of an inflammatory condition, may simplify existing protocols.

Two much larger clinical populations reported in the meantime further suggestions for AAT reference ranges [211,212]. While their lower limits were comparable to our results, their upper limits were substantially higher [213]. This difference does not seem to originate from the quantification methods applied since one study used immunoturbidimetric assays as we did in our analysis [212], and the other used nephelometry [211], a method which is considered to produce equal results. Phenotyping was used in both studies to determine deficiency alleles. Some rare genotypes could therefore not be determined compared to the exon-sequencing method which we applied in our study, but this at most marginally affected the lower boundary of the reference ranges. Impact of sex or age was judged as minor as we also did in our analysis. The reason for the elevated upper limits was most likely the fact that the results were not derived from general population samples. Genotype distributions were in fact enriched with deficiency alleles and deviated substantially from Hardy-Weinberg equilibrium. Since many subjects of the clinical populations may show presence of low inflammatory conditions, the cut-off mentioned in the previous paragraph may be too low in patients with clinical symptoms.

7.1.5 Novel Candidate SNPs for the Longitudinal Air Pollution-Lung Function Association

Benefits from cleaner air are not equally distributed among people, but are influenced by the individual genetic profile. As shown before, *SERPINA1* deficiency variants did not interact with air pollution decline in terms of lung function loss in SAPALDIA. Clear evidence for interaction with air pollution is also missing for polymorphisms in oxidative stress genes [119]. The strategy chosen here, namely testing GWAS-identified SNPs for interacting effects with environmental exposures, was successful for breast cancer [214]. However, we could only find modest modifications of the effect of reduced PM_{10} on longitudinal lung function by SNPs selected for their strong association with lung function (chapter 6.1). Interactions did not remain statistically significant after adjustment for multiple testing. Knowledge about the functionality of the genes in which those SNPs are localised was absent and could have been improved to a certain degree if compelling evidence for interaction had been present. The set of SNPs did not interact with smoking [131]. Moreover, they did not even show association with longitudinal lung function in a subset of SAPALDIA [133]. Their robust association with cross-sectional lung function might therefore be explained by influencing lung function growth during childhood and adolescence rather than lung function decline later in life. It remains to be determined if gene-environment interactions are more likely observable for SNPs not reaching genome-wide significance.

We should also keep the explanatory potential of epigenetics in mind. Short-term exposure to air pollution could already be associated with differences in DNA methylation [215]. The clarification of the mechanisms for the different susceptibilities to air pollution with regard to lung function remains of particular interest. Air pollution is unavoidable and a growing public health problem in densely-populated urban centres caused by production facilities, heating sources and traffic. Up to now, we do not know an exposure threshold for particulate matter below which adverse health effects can be avoided.

7.1.6 Bridging the Gap between Obesity and Asthma: a Role for Altered Cell Division?

Potential mechanisms how obesity could cause asthma include changes in inflammatory marker concentrations or in molecules secreted from adipose tissue (adipokines) [216]. This

may lead to remodelling processes in the airways, and genes influencing cell cycle control in cells of the respective tissues seemed therefore promising candidates for effect modification. Our results suggest that the *CCND1* gene might play a role in the obesity-asthma association (chapter 4.1).

This first project of the PhD work has several limitations reflecting in part the methodological progress in genetic epidemiology since then. We tested only three candidate genes covered by one SNP each due to genotyping cost constraints. Although these SNPs were suggestively functional since previously associated with a few cancer types in case-control studies, such a meagre selection of SNPs would hardly be envisaged today. Functional annotation of the SNPs (e.g. by eQTL data) or a completer SNP coverage of the genes would be mandatory to draw firm conclusions from the results. Cellular pathways shown to be important in asthma [217] would lead to a less arbitrary gene selection. Generally speaking, such a finding remains isolated and earns a lot of scepticism if not replicated in an independent study sample or, even better, in a hypothesis-free approach. A genome-wide search for SNPs associated with BMI in individuals with and without asthma did not reveal any loci which were specific for asthmatics apart from *DENND1B*, a gene functionally related to clathrin-mediated endocytosis, which was only associated in asthmatic children [218].

Since asthma is a binary outcome, we could have used a case-only design to assess the gene-environment interaction with improved statistical power. But two disadvantages would have applied. First, the estimation of the separate effect of the genetic and the environmental factor is not possible. And second, since the independence of the genetic and the environmental factor is a central prerequisite, the design is not suitable if the environmental factor is a type of behaviour like obesity (or smoking).

7.1.7 Candidate Gene-Environment Interaction, a Cautionary Note

Four of the six projects of this PhD work were based on gene-environment interactions. In all of them, we found some evidence of the presence of such interactions. As for the interactions with AAT deficiency variants, there was a lot of prior knowledge available which argues in favour of the presence of such interactions. The involvement of environmental risk factors like smoking and other inflammatory triggers seems in that case a

much more promising approach in order to identify associations than simply assessing genetic main effects. For the other interactions, findings must be treated with more caution. While results from candidate gene association studies do not cause much interest in the research community without a successful replication in independent study samples (see 1.1.4), the scepticism towards non-replicated candidate gene-environment interaction studies is even higher. The general lack of compelling gene-environment interaction results is one reason for this and is explainable by the challenging sample size requirements as well as the usual non-availability of environmental factors in a uniform manner across studies (see 1.1.8). But owing to this latter issue, it is nonetheless important that single studies generate, test and publish interaction hypotheses. These results may later stimulate similar research in much larger and better characterized cohorts and associated biobanks, which will likely be available in the future (see 7.2.5).

7.2 Outlook on the Epidemiological Research of Complex Diseases

By understanding the impact and interplay of the different risk factors, large progress in prevention, risk prediction, diagnostics and treatment of complex diseases could be envisaged. We will mention here current trends from a genetic, environmental and combined perspective and close this chapter by listing some open questions regarding AAT, the main focus of this thesis.

7.2.1 The Genetic Perspective

Well-conducted candidate gene association studies may still contribute to the gain of knowledge in the future. In fact, a recent comparison of statistically significant cancer associations between meta-analyses of candidate gene association studies and GWAS found similar effect sizes, but nevertheless little overlap in statistically significant results [219], a finding which was also shown for other outcomes [220,221]. This suggests that the SNP selection procedure in candidate gene association studies was often suboptimal in the past. Common variation in regulatory regions seems more disease-relevant than common variation in coding regions. Candidate genes should therefore be better represented by non-exonic SNPs with functional annotation.

The GWAS design will continue its success story of the past seven years by further enabling collaborations of many large consortia and providing knowledge about mechanisms in a large number of NCDs and medically-relevant traits [10]. Although the genetic and environmental context is in most GWAS not taken into consideration, the recent developments (see 1.1.6) may still be able to reveal many more robustly associated genetic variants. It was claimed that by increasing sample size and using most modern chip designs, more than five times as many GWAS hits as are currently known for chronic diseases could be found whereby sample size was alleged to have the far bigger influence than chip design [21]. The impact of meta-analyses may still grow, but the random-effect design should be promoted to better estimate the magnitude of the effect size [20]. Projects like the UK10K will increase the available reference panels substantially and will lead to high imputation quality down to a MAF of 0.5% or even less [222]. The increase of sample size is however not an infinite process. For some traits like human height or lung function, the amount of participants in current GWAS meta-analyses reaches already six-digit numbers. Alternative ideas are in need. For instance, the newest GWAS on lung function concentrated on a founder population which provided increased statistical power due to the homogenous genetics and environment [223]. This way, some of the previously known associations were confirmed, and by analysing several SNPs simultaneously in the same model [207], a few novel findings were reported.

While a few of the follow-up methods mentioned in 1.1.7 are meanwhile regularly applied, others are still in its infancy and possess high potential to detect more disease-associated variants. WES has already proved to be valuable for the discovery of rare variants in NCDs [224]. It is however only a stopover to WGS, for which a first association analysis with a complex quantitative trait has just been published [225]. WGS still faces important challenges. Sequencing costs are still high, namely between US\$ 5000 and 10 000 per genome and have currently for the first time stopped going down (Figure 4, [226]).
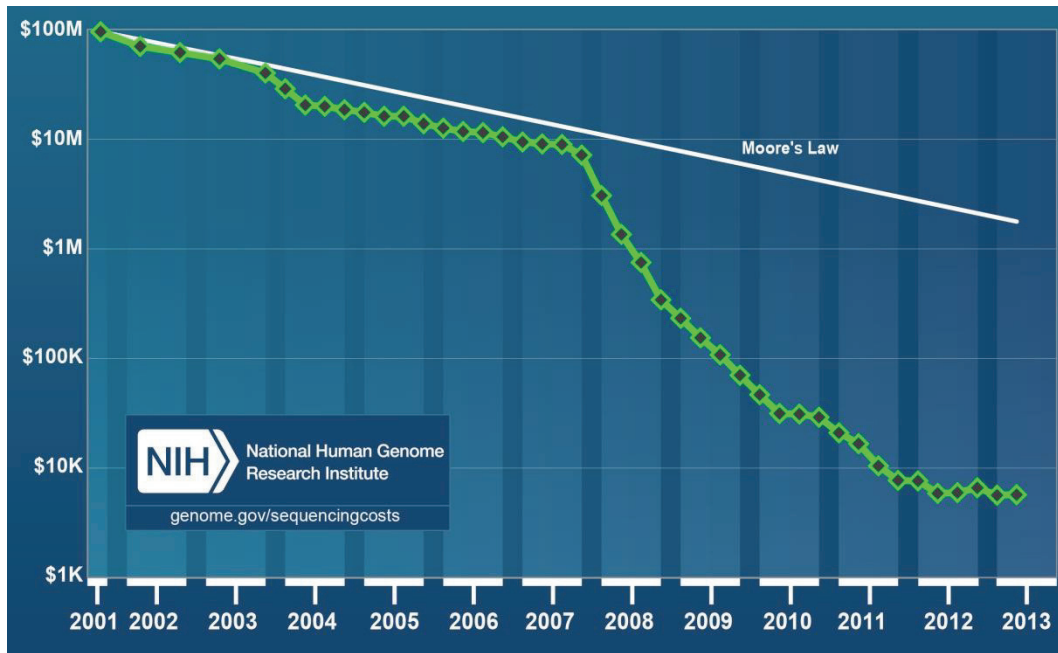
Figure 4. The development of the sequencing costs per human genome.

Large sample sizes are needed to have sufficient power to detect disease associations with rare variants. The validity of such association studies is not yet established. The complementary use of family-based designs may help in the interpretation of results. Aggregating variants at the gene level improves power, but challenges the use of meta-analyses. In this active area of research, methods using study-level specific gene-level summary statistics have been proposed [227]. Further unsolved issues include the correction for population stratification and the determination of the genome-wide significance threshold.

Progress will result in detecting further disease-associated variants, probably with mostly small effect sizes. Some of the missing heritability might be explained, but it is well conceivable that due to non-additive components or generally inflated heritability estimates, current and future approaches will not reveal many more variants than currently known [39,40]. More importantly, they should help identify the causal variants underlying the signals already identified.

7.2.2 Translation into Clinics

The most advanced area in terms of clinical utility of GWAS findings is pharmacogenomics. By measuring differences in drug side effects or drug clearance in

affected patients, GWAS revealed several genetic variants with sometimes large effect sizes [202,228-232]. It can be speculated that harmful variants with large effect sizes could be more easily determined in this research field due to the absence of selection pressure until recently. A further advantage of this field is the use of RCT in order to confirm GWAS findings. Translation into clinics has already taken place, e.g. when starting antiretroviral therapy, a screening for a gene variant which provokes a strong side effect against abacavir may be conducted [233].

With regard to complex diseases, the situation is more challenging. GWAS results have not played any clinical role yet, a fact which makes some people criticize the entire method. However, all GWAS findings are comparably new, and clinical applicability can certainly not be expected in such few years. The main problem towards a personalized medicine including prediction of risk, effective prevention and optimal treatment are the small effect sizes of the identified variants. Further issues include their incomplete penetrance, the moderate heritability they are able to explain and the lack of knowledge of causality. Since it is not clear how many risk loci should be included in a test, the non-accessibility of many GWAS results, especially of those below the genome-wide significance threshold, is also of concern.

GWAS have already led to the discovery of novel disease mechanisms and new therapies may evolve from that. In terms of prevention, individual genetic prediction tests may become useful in the future, e.g. in order to detect the most susceptible subgroups which may profit most from interventions. The accuracy of such a test can be measured by its sensitivity, specificity and the associated area under the receiver operating characteristic curve (AUC). Genetic tests have a few useful properties like their time-independency and the easy non-invasive collection of DNA, e.g. from buccal cells. The success of a genetic test for a certain disease depends largely on the heritability of the disease which we can currently explain. There are several examples in which SNPs selected from results of candidate gene association studies or GWAS were of no use to predict an outcome (e.g. for alcohol dependence [234] or childhood asthma [235]). There is however also evidence that genetic tests for traits or diseases for which very large GWAS exist show better performance in terms of AUC [236]. For instance, tests based on current GWAS on type 2 diabetes perform much better in terms of AUC than those based on previous GWAS [237,238]. Mild increases in AUC can have a large influence at the population level by

allocation of public health resources for prevention in risk groups. For some diseases, the use of formerly identified SNP associations resulted in more accurate tests than family history [239]. And it was suggested that the combined score over all genomic markers could substantially improve disease prediction [240]. A disease for which genetic prediction works well is type 1 diabetes. Its genetic components are comparatively well described (i.e. the missing heritability is small), and the AUC reaches 0.9 [236]. But since this disease is rare, the positive predictive value, i.e. the number of correct positive predictions compared to all positive predictions, may still be low. By applying type 1 diabetes prediction tests only to persons with a familial risk, the positive predictive value could be greatly improved. However, testing all persons belonging to a risk group (screening) is only useful when preventive strategies are available. In the case of type 1 diabetes, they are not yet available, but progress is being made [241].

Polygenic risk profiles will certainly find their way into preventive medicine complementing environmental risk profiles, traditional clinical profiles and profiles based on family history. The identification of concordance and discordance among such profiles may confer important information for a personalized health action plan [242]. It is currently not clear if genetic prediction tests based on SNP genotyping may at some stage be replaced by WGS. The advantage of WGS would be that no genetic region has to be specified prior to sequencing. The power of WGS has already been shown by its use in the detection of monogenic disorders [243]. Moreover, there are reports in which a genetic disorder could be determined and, simultaneously, the same genomic information was used to decide on the best therapy [244]. Investments into WGS are currently huge. The Prime Minister of England has announced a 100k Genome Project in which up to 100 000 patients are to be sequenced in the next five years. This data could then be linked to medical records. It was suggested that consent to genetic research could be given in a similar way as for organ donation. The Faroe Islands with a population of about 50 000 inhabitants plan to be the first country in the world which sequences the genome of every citizen and link it then to full digital health records.

But this boost of WGS comes at a potentially high price. There is an increasing chance to detect incidental findings with a large influence on health. The European Society of Human Genetics recommended therefore that targeted sequencing limited to genome regions linked to the clinical problem should still be encouraged as opposed to WES and WGS [245].

Furthermore, WGS should only be an option if an unsolved clinical problem is present. Incidental findings should be reported if indicative of a serious health problem and if prevention or treatment is available. This should be weighted larger than the patient's right not to know. Incidental findings may also have implications for the children or for family planning. Further challenges of WGS include the limited data storage capacity in many hospitals. It seems crucial that informed consent for diagnosis, research, disclosure and data storage is addressed separately. A new global alliance is setting out to ensure on the one hand the provision of suitable protection for the autonomy and privacy of the people who have contributed their data and on the other hand that the benefits of genomics sequencing for health are realised as efficiently as possible [246].

7.2.3 Direct-to-Consumer Genetic Testing

Several companies, usually operating from the USA, offer meanwhile private genetic testing direct to consumers (DTC) for a modest amount of money. The concept is shown in Figure 5. Saliva samples get genome-wide analysed and relative risks for a wide range of diseases are calculated. Additionally, carrier statuses for several recessive disorders are determined. All this information is provided online through a password-protected user account.
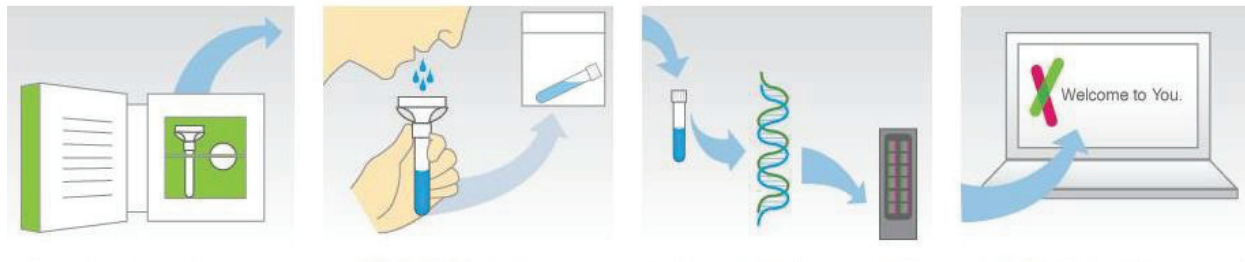
Figure 5. The concept of direct-to-consumer genetic testing. *Source: 23andMe*.

Due to the lack of counselling, DTC genetic testing is considered especially inappropriate for individuals with disease symptoms, for high-penetrance serious disorders, for prenatal screening, testing of minors and pharmacogenetic testing [247]. The legislative situation concerning such services is very heterogeneous among European countries [248]. In Switzerland as well as in several other countries, DTC genetic testing is forbidden. In the USA, there is a trend towards stricter regulations of DTC genetic tests as the Food and Drug Administration have asked the companies offering such services to seek regulatory approval for their tests.

The idea behind such services is often advertised by four P's: predictive, preventive, personalized and participatory. A DTC genetic test with own DNA issued by the company "23andMe", one of the largest providers of such services, pointed to several weaknesses though. The calculations for many of the traits and diseases were only based on very few SNPs and did insufficiently represent the current state of knowledge. Hence, the predictive capacity was for most diseases low if not inexistent, and claims seemed overstated and could potentially provoke anxiety or confusion. Results were provided independent of available preventive measures (e.g. results for Alzheimer-associated variants were conveyed). The genetic data of all participants are also included in research studies. This participatory idea must be acknowledged since it offers the appealing potential to build up huge data sets in a minimal amount of time. Participants are regularly invited to provide topic-specific health data, and social media are used to spread these services. "23andMe" was for instance able to successfully replicate many GWAS findings from the literature in their online communities, but given the large statistical power they had, the agreement was lower than expected [249]. This might have to do with the shortcomings in the data-collection process. The fact that only self-reported information by volunteers is considered may result in selection bias (e.g. volunteers are better educated regarding health than the general population) or reporting bias (e.g. individuals interested in a certain disease may recall relevant exposures differently than others) [250]. The sheer mass of participants makes such research nevertheless valuable.

Receiving information about a considerably elevated risk of getting a severe disease in the absence of genetic counselling may have important implications on a person's well-being. Web-based networks like "PatientsLikeMe" try to facilitate the sharing of benefits, but they can neither replace the opinion by an expert nor take the personality profile of the proband into consideration. A recent survey including European clinical geneticists showed that almost all participants were opposed to any form of DTC genetic testing for medical conditions that were either serious or not preventable or treatable [251]. A large majority did not support DTC genetic tests for highly predictive conditions, but about 40% did not object to carrier testing. Most said they would be willing to see patients who were concerned about DTC genetic testing results, and almost half of them have already done so. There was some disagreement on the question if genetic counselling is indispensable for tests without potentially serious consequences.

While only preliminary knowledge is present about the behavioural consequences of genetic testing [252], it is well established that the genetic illiteracy in the average population is high [253]. But such literacy would be important for the decision whether or not to disclose genetic information to the patient. Results must be explained in an understandable way which includes information about the high degree of uncertainty in quantifying risk prediction for polygenic diseases. Carrier status of potentially dangerous and often high-penetrance risk alleles is on the other hand information which can be determined very reliably and which may have stronger implications for health and family planning. The decision for disclosure of such information must carefully balance the patient's autonomy, psychological welfare and health. This applies also to the disclosure of incidental findings outside the scope of offered services which may become important if such services switch to WGS. It is furthermore crucial that companies inform and get the patient's separate consent about the storage of data, the possible use for research and the sharing of the data [250]. The prohibition of DTC genetic testing is probably not a reasonable permanent solution due to the globally easy accessibility of such tests. In fact, there are efforts in Switzerland to allow DTC genetic testing for providers who inform adequately about purpose and risks of such tests. But once legal, health insurances are expected to profit from this situation and complete their application forms with questions about the presence of such genetic tests. In that case, persons who want to carry out DTC genetic testing just out of interest must be aware that they may potentially regret this decision.

Taking together, the relevance of personalized medicine is increasing in the medical as well as in the private sector. While genetic tests for some rare high-penetrance diseases are very reliable, this is not (yet) the case for most complex diseases. The limits are clearly seen by the fact that genetically identical twins rarely develop the same complex diseases [254,255]. The environment as well as related genetic and epigenetic modifications in somatic cells of several tissues soften the image that the genetics is an unchangeable and highly predictive parameter of human health [256]. It is important not to forget that chronic diseases are to a large part environmentally driven and often preventable by modifying the environment. Nevertheless, the young discipline of public health genomics features a full agenda [257]. Social and environmental models, public health professionals usually work with, should be widened by the integration of genetic knowledge. Interventions may be increasingly evaluated according to genetically-defined population strata. But most importantly, the genetic literacy of medical practitioners, who may see more patients coming with DTC

genetic test results to them, as well as of the public should be increased. RCTs are necessary to compare outcomes between patients who received counselling without genetic information and those who additionally received genetic information [258].

7.2.4   The Environmental Perspective

Environmental factors are much harder to describe in their completeness than genetic factors. Namely, for any specific place and time, not all relevant exposures are known. Moreover, measurement accuracy depends strongly on the type of exposure, some exposures fluctuate strongly over time, and biomarkers are not readily available for most environmental exposures. Nevertheless, the inclusion of environmental factors in prediction tests reaches often much higher AUC than the use of genetic factors (e.g. in risk scores for cardiovascular outcomes, [259]). Since many of these risk factors can be modified, they should stay in the centre of prevention. In order to find novel risk factors independent on previous knowledge, agnostic procedures analogous to GWAS (so-called environment-wide association studies) have been carried out (e.g. [260]). Challenges include reverse causation (especially if there are no longitudinal data available), confounding and dependency of exposures on one another (see 1.1.2). Similar to LD-considerations in GWAS, exposure correlations can be used to refine signals.

An attempt to evaluate the comparative importance of various diseases and risk factors in the population is used in the Global Burden of Disease Study [1]. Importance of risk factors was estimated by comparing the effects of each exposure with a theoretical minimum population distribution of the exposure. Shortcomings include the assumptions that risk factors act independently of one another and that exposures are uncorrelated across diseases. Caution should therefore be applied when using the ranking of the 67 risk factors to guide policy in certain regions [261]. This statement is intensified by several additional issues. First, the ranking does not consider the feasibility or cost-effectiveness of interventions. Second, several risk factors are not included in the study (e.g. unsafe sex, poor hygiene, education or noise) due to a trade-off being made between high evidence of causality and data availability. Country-specific cohort studies which include the assessment of a wide range of risk factors would be needed to combat the widespread lack of data and thus the dependency on complex and error-prone estimation techniques in certain regions. Finally, it is worth mentioning that NCD research has been carried out mainly in high-income

countries. This is in contradiction with the fact that low- and middle-income countries (LMICs) account for 80% of NCD deaths [262], a percentage expected to even rise in future [263]. The transfer of knowledge from high-income countries would only partially solve the problem due to other distributions of exposures and confounders in LMIC. High-income countries may equally benefit from a boost in the allocation of resources to NCD research in LMIC [264], but it is ethically important that LMIC must also benefit from the implementation of knowledge thereafter. Stronger health information systems involving the registration of births and deaths are a much-needed first step towards better health data in poorer areas of the world [265].

An interesting aspect is the complete absence of genetics in the Global Burden of Disease Study. This is in part explainable by the fact that, although the joint influence of genetics on complex diseases is often well investigated and large, the single causal variants are usually not known with high certainty, especially not across global populations. It does however not seem essential to aim at expressing the risks of genetic variants in terms of DALYs. While genetics is crucial to determine disease mechanisms and defining susceptible subgroups in the population, we should not forget that the environment is the component we can influence most easily. The genetics is not a factor which prevention can act on.

There is hope that sensor and mobile technology will lead to a much more accurate and comprehensive measurement of environmental exposures, behaviours and lifestyle factors in the near future. Such a development could boost the environmental research in a similar way like observed for the genetic research owing to the new technologies in the past decade.

7.2.5   A Combined Perspective

In terms of assessing gene-environment interactions, we may expect that, analogous to the genetic main effect literature, agnostic study designs like GWIS become the preferable approach. Such studies may elucidate mechanisms for genetic associations with an outcome, potentially contribute to reduce the missing heritability and identify susceptible subgroups in the population. Some GWIS findings have been reported up to now [94,132,266-268], but the methodology is less standardized than in GWAS. Results with recently suggested methods seem to mainly reflect the marginal effects of the genetic variants [269]. In fact, the scarcity of reliable examples let to the suggestion that genes involved in interactions may

better be detected through their marginal effects [270]. Large sample sizes are a prerequisite in GWIS for being sufficiently powered (see 1.1.8). This can be achieved by the use of large collections of environmental data or, alternatively, by the use of meta-analyses of smaller data collections. The latter increases the risk of heterogeneity in exposure assessment or of differences in the assessment of confounders. Potential higher order interactions are difficult to consider since they strongly increase the burden of multiple testing.

Another approach of investigating associations between environmental factors and diseases in dependence of the genetics may lie in the investigation of intermediate processes (e.g. changes in gene expression or epigenetic modifications). The key task to clarify these processes is the setting-up of comprehensive biobanks. There has been considerable investment in biobanks recently, and internationally harmonized efforts are crucial to keep comparability high. Biobank research comprises scientific, infrastructural and ethical aspects. Focusing on the scientific aspect, a biobank should ideally include samples of many tissues and several time points since epigenetic and expression profiles vary tissue- and time-dependent. An important issue is the identification of biomarkers as proxies for environmental exposures. The collection of samples before and after the presence of certain exposures is an intriguing approach to investigate short-term exposure effects. A recent "omics"-approach showed that genomic, transcriptomic, proteomic and metabolomic information of the same individual over several months can reveal important pathways and allele-specific expressions during disease periods and can be used to predict, diagnose and even treat disease-related conditions [271]. This represents a future scenario how health monitoring and personalized medicine could work together. An additional and even less understood component is the epigenome. Epigenetic modifications were observed to be associated with many complex diseases including asthma [272]. While several exposures were also associated with genome-wide or gene-specific methylation, the causality in the epigenetic association with the disease remains an unsolved issue.

7.2.6   The Potential Contribution of SAPALDIA

SAPALDIA is the only longitudinal and population-based study in Switzerland which covers all language and cultural regions. Started as an air pollution and lung health study, researchers later widened the focus to several chronic diseases. The added biobank in the second survey was a key achievement for establishing many collaborations with other

European cohorts. Since the third survey in 2011, the biobank has contained longitudinal serum and DNA samples. This would potentially enable investigators to look into the metabolome and epigenome over the course of ten years. Another new intriguing opportunity is to investigate transcriptomics, metabolomics or epigenomics as a function of short-term exposure. The number of participants with available genome-wide data was recently enlarged to approximately 4000. The availability of almost one million genotyped SNPs and many more imputed ones making use of the newest 1000 genomes reference panels will generate new collaborations and foster many existing ones. Of note is that more than half of the study participants meanwhile attained an age of over 65 years, which shifts the SAPALDIA focus to a an elderly population. This is an opportunity to include more chronic diseases which get particularly prevalent at an older age like for instance cognitive diseases.

With a sample size of roughly 5000 individuals, it is clear that the establishment of collaborations with all the accompanying difficulties in comparing results across different studies remains a crucial task. A large-scale national cohort in Switzerland would facilitate this situation and allow researchers to investigate NCDs in a much wider and more flexible framework. Other European countries like Germany or Great Britain invest in such mega-cohorts and may serve as examples how to best make use of the profound pre-existing expertise with population-based cohorts present in Switzerland.

7.2.7   Future Research in Relation to AAT

After this general outlook on the future research on the aetiology of NCDs, a summary of open questions arising from the PhD projects regarding AAT will complete this thesis.

We observed in SAPALDIA that subgroups of heterozygous PiZ carriers, namely persistent smokers, obese persons and those highly exposed to vapours, gas, dusts and fumes, are of increased risk for elevated lung function decline. Replication in other cohorts is a first step to consolidate these findings. There is evidence that smoking and high occupational exposure are also important for the differing susceptibility of PiZZ individuals to adverse pulmonary health [164,273], findings which support our results. The effects of obesity in combination with severe AAT deficiency have not yet been evaluated, and such a project could be envisaged with data from an AAT registry including ideally index (i.e. those

presenting symptoms of adverse lung (or liver) health) and non-index cases. Unlike in studies examining PiZZ individuals [149], longitudinal air pollution was not observed to interact with PiMZ genotype with respect to lung function decline in SAPALDIA. However, we must consider that exposure levels are comparably low in Switzerland. It would be interesting to investigate this issue in areas with high or strongly rising levels of air pollution like some urban regions in India or China. Projects including indoor air pollution may strengthen the hypothesis that individuals with intermediate AAT deficiency have less capacity to combat the proteolytic effects of inflammatory triggers. However, we must take into consideration that those regions in which indoor air pollution is a highly relevant risk factor for adverse pulmonary health are largely free of the PiZ allele [274].

It is currently not clear if heterozygous PiZ carriers exposed to inflammatory triggers benefit equally from a reduction of the inflammatory trigger. A randomised smoking cessation study in smoking PiMZ individuals may clarify this issue. A better strategy would of course be the counselling of PiMZ subjects before the exposure exerts its effect. Studies may assess which type of counselling is most effective. PiMZ individuals could be randomised to counselling including or excluding information about AAT genetics. It could also be assessed in a similar way if regular counselling (e.g. on a yearly basis) is superior to one-time counselling. The possibility of recruiting large numbers of PiMZ individuals also depends on complete information on PiZZ subjects in AAT registries. National AAT registries as well as the Alpha One International Registry, which collects data of the national registries of nearly 20 countries, aim to capture severely deficient individuals as completely as possible with the goal of encouraging research and promoting knowledge. However, by roughly comparing numbers of registered PiZZ individuals with those estimated based on population-based studies, severe AAT deficiency appears a strongly underrecognised genetic disorder with merely about 10% registered individuals.

While one reason is the absence of clinical symptoms in a large group of deficiency carriers, the often long time-lag between the appearance of first clinical symptoms and the diagnosis of an AAT deficiency is another aspect. This is of high importance since the deficiency diagnosis may additionally motivate to give up smoking and may result in an early start of augmentation therapy (if available) and hence a probably better long-term prognosis. Moreover, family members could be tested and counselled at an earlier time point, which is generally the best method to detect early disease and prevent from more serious disease.

Reasons for the time-lag in the magnitude of seven to eight years are the inadequate awareness of healthcare providers and the failure to implement available guidelines [275]. The strongly heterogeneous opinion on AAT augmentation therapy among experts may also represent an important point in the decision in favour or against testing [276]. Educational campaigns for primary care physicians or pulmonologists, distribution of free test kits to measure AAT serum level to physicians and even dried blood spot kits for a home-based test with the offer of free genetic counselling over the telephone were measures of a testing programme which also aimed at detecting heterozygous PiZ carriers [277]. Services offering DTC genetic testing like "23andMe" also routinely assess the PiS and PiZ deficiency alleles and could potentially provide carriers with important information about the purpose of AAT registries. The search for non-index cases is namely an important part of testing programmes since the counselling of these symptom-free individuals may be much more promising with regard to the prevention of severe pulmonary disease.

To whom is it recommended to undergo genetic testing? Generally, this depends on the prevalence of the genetic syndrome, the prevalence of the associated diseases, the disease burden, the availability of interventions, the chance that an intervention is accepted, the accuracy of the test, the costs as well as on psychological and ethical aspects. The most rigorous approach would be neonatal screening. Experience with such an implementation comes from Sweden where all newborns between 1972 and 1974 were screened for AAT deficiency. More than half of the 127 PiZZ-diagnosed babies had abnormal liver tests at the age of three months [140], but in most cases values normalised later during childhood. This is in line with the fact that less than 1% of all PiZZ children need liver transplantation in early childhood [278]. In a follow-up study of the neonatal screening project, the families from the diagnosed PiZZ babies got ample information about the risks for their babies. Although this initiated negative emotional reactions especially in mothers, the majority of parents were positive towards having received the diagnosis with extensive counselling. While parents of PiZZ children did not smoke less than those of PiMM carriers, the PiZZ carriers themselves smoked substantially less as adolescents than their PiMM peers [279]. Adverse psychological effects in the adolescents could not be found [280]. Although these results from Sweden seem generally supportive of a population-wide screening before the age of adolescence, there is currently no country which conducts or recommends population-wide screening for AAT deficiency. While the genetic syndrome is indeed relatively frequent (e.g. more frequent than cystic fibrosis or phenylketonuria), the

penetrance of the associated diseases is not very high, especially if risk factors like smoking are avoided. The second major reason why neonatal screening is nowhere implemented is probably the absence of a directly beneficial intervention during childhood (apart from the avoidance of exposure to ETS). A screening close to the age when smoking may start would possibly be more beneficial and would allow the child to take part in the screening decision. Uncertainties include the participation rate of a voluntary AAT screening and the efficacy of anti-smoking advice if the information is given to school-age children [281].

The screening of asymptomatic individuals who are at elevated risk for adverse pulmonary health is a second strategy. Such groups include persons with a family history of unexplained liver disease, early-onset emphysema or AAT deficiency. The guidelines of the American Thoracic Society and the European Respiratory Society recommend testing especially to siblings of PiZZ carriers since they have a 25% chance of having the same genotype [173]. The first step in the determination process does however not yet involve genetic testing, but measuring levels of circulating AAT (chapter 5.1). A volunteer screening of smokers or people working in jobs with high exposure to vapours, gas, dusts and fumes could also be conceived. In order to assess the benefit of offering such a service to a substantial proportion of the general population, one would have to carry out a RCT in which the testing offer is randomly assigned to the exposed population. Since health-relevant outcomes like lung function may show effects only years or even decades later, costs for such a study would be prohibitively high. It is also important to emphasize that prevention from inflammatory triggers is essential for all individuals. Even if genetic subgroups are found to be more susceptible, this does not imply that efforts for subgroups less susceptible are of lower importance. In that respect, the offer of a genetic test for subjects highly exposed at work should by no means weaken efforts to generally lower occupational exposure.

A different situation applies to symptomatic individuals. Subjects with unexplained liver disease, with early-onset emphysema or COPD, or with airflow obstruction in the absence of risk factors should be screened for AAT deficiency, at least in regions where AAT deficiency is prevalent like in Europe and the USA [173]. Interventions for those persons consist of general measures for the treatment of COPD (i.e. reductions of exposure, provision of oxygen supply and use of bronchodilators, oral corticosteroids and antibiotics) as well as potentially of intravenous AAT augmentation therapy. The prescription of the

expensive therapy is subject to several regulations. While a few countries like the USA aim at generally prescribing replacement therapy to all severely deficient individuals, others like Switzerland allow it only for individuals who are severely obstructive and have quit smoking. A few other European countries like Great Britain or the Netherlands do not offer substitution therapy at all.

Considering the uncertain profit from AAT replacement therapy in PiZZ individuals, there is at this point no indication for assessing its use in PiZ heterozygotes, not even in those additionally exposed to other risk factors. Experimental research is first needed to solve the question why the establishment of normal AAT blood levels does not convincingly result in a normalised rate of lung function loss. The role of locally produced PiZ polymers in the lung must be clarified. A difficulty in using mouse models is the deviation of the human *SERPINA1* locus from the murine orthologue. Gene duplication led to several paralogues in mice, some of which seem to be essential during embryogenesis [282].

Several new approaches are being developed in order to have alternatives to AAT substitution therapy. Viral vectors and liposomes have been tested as vehicles for gene therapy. Single administration would be the main advantage as opposed to the weekly administered replacement therapy, but a few challenges still need to be resolved. On the one hand, only slightly and short-term elevated AAT levels could be observed after vector injection. On the other hand, PiZ formation and polymerisation in hepatocytes remain unchanged by current gene-therapy approaches [283]. Aerosols instead of intravenous application of replacement therapy have been proposed, but delivering large proteins by inhalation is challenging. Production of human AAT in transgenic animals would substantially lower the costs as opposed to the purification of human plasma, but did so far not result in a molecule tolerated by humans [284]. Finally, a strategy to stimulate degradation pathways in order to clear hepatocytes from AAT polymers was suggested [285].

Our projects concentrated on environmental factors modifying the influence of *SERPINA1* deficiency alleles on lung function decline. However, there may also exist important genetic modifiers. In prediction models, the inclusion of age, sex, packyears of smoking, bronchodilator responsiveness, chronic bronchitis and index status explained more than 50% of the FEV_1 variance in severely AAT deficient subjects. SNPs of four candidate genes previously associated with pulmonary function in PiZZ individuals and their interaction

with smoking could not contribute to a more accurate model [286]. SNPs in *CHRNA3*, formerly associated with COPD, have been found to marginally modify airflow obstruction in severely deficient persons [287]. We expect more genetic co-factors to determine lung function in those people, and such information could be used in the decision process in favour or against AAT substitution therapy. A GWAS on lung function decline involving only PiZZ individuals would be an interesting approach to find such variation, but only a collaboration of different AAT national registries like the Alpha One International Registry would lead to sufficient samples to conceive such a project.

The underlying genetics of AAT serum level outside the *SERPINA1* coding region could not be clarified in this PhD work, but is unlikely to play an important role in influencing pulmonary health as even the PiS variant, which considerably lowers AAT serum level, could so far not be associated with pulmonary health. However, there is growing evidence that AAT is more than just a protease inhibitor of neutrophil elastase [284]. Antibacterial and -viral properties have been observed, and its anti-inflammatory activity affects a wide range of immune cells such as T lymphocytes and dendritic cells. Novel experiments in elastase-deficient mice as well as with a recombinant form of AAT without anti-elastase activity provided evidence that the anti-inflammatory and immunomodulatory properties of AAT are independent of elastase inhibition [288]. AAT may thus have the potential to achieve clinical value for a variety of diseases and is currently tested as a therapeutic agent for cystic fibrosis and type 1 diabetes. Its pleiotropic potential is further emphasized by its suggested role in CVDs. *SERPINA1* variation was associated with different measures of atherosclerosis [28,289,290] and increased aortic stiffness [291]. AAT appears essential for the integrity of the connective tissue in blood vessels, and damage to the elastic fibres of the arterial wall is assumed in severely deficient individuals [292]. Experimental studies have further shown that AAT occurs in human atherosclerotic lesions [293]. Given this wide range of effects, it is crucial to better understand the genetics of AAT expression beyond the large effects of a few exonic variants leading to misfolded proteins. The meanwhile much larger SAPALDIA sample that is genome-wide analysed and the better reference panels for imputation would enhance the statistical power to detect loci which influence the serum level more moderately than currently known variants in a GWAS replication project. Pathway analysis could then be useful to test if GWAS results are enriched in inflammatory pathways or in cholesterol or triglyceride pathways to which *SERPINA1* was recently found related [28].

Finally, expression of inflammatory mediators may be programmed by epigenetic mechanisms. While *SERPINA1* methylation was observed to be associated with lung function and COPD [117], its association with serum level has not been investigated so far. The data of SAPALDIA would allow an epigenome-wide association study with AAT serum level to potentially find further genes relevant to the *SERPINA1* gene expression. More specifically, the *SERPINA1* methylation pattern could be compared between subgroups with and without deficiency alleles and high exposure to inflammatory stimuli, potentially clarifying the observed association between *SERPINA1* methylation and lung function.

8 References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, et al. (2012) Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380: 2197-2223.

2. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, et al. (2012) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380: 2224-2260.

3. Liu CY, Maity A, Lin X, Wright RO, Christiani DC (2012) Design and analysis issues in gene and environment studies. Environ Health 11: 93.

4. Ebrahim S, Davey Smith G (2008) Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet 123: 15-33.

5. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet 9: 255-266.

6. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4: e1000008.

7. Powell JE, Henders AK, McRae AF, Kim J, Hemani G, et al. (2013) Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. PLoS Genet 9: e1003502.

8. Tenesa A, Haley CS (2013) The heritability of human disease: estimation, uses and abuses. Nat Rev Genet 14: 139-149.

9. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76-82.

10. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90: 7-24.

11. So HC, Gui AH, Cherny SS, Sham PC (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol 35: 310-317.

12. Zaman MJ, Bhopal RS (2013) New answers to three questions on the epidemic of coronary mortality in south Asians: incidence or case fatality? Biology or environment? Will the next generation be affected? Heart 99: 154-158.

13. Kavvoura FK, Ioannidis JP (2008) Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 123: 1-14.

14. Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470: 187-197.

15. International HapMap C (2005) A haplotype map of the human genome. Nature 437: 1299-1320.

16. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.

17. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38: 209-213.

18. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2: e190.

19. Hindorff LA, MacArthur J, Morales J, Junkins HA, Hall PN, et al. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed: 01 September 2013.

20. Evangelou E, Ioannidis JP (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14: 379-389.

21. Lindquist KJ, Jorgenson E, Hoffmann TJ, Witte JS (2013) The impact of improved microarray coverage and larger sample sizes on future genome-wide association studies. Genet Epidemiol 37: 383-392.
22. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, et al. (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 8: e1002793.
23. Teo YY, Small KS, Kwiatkowski DP (2010) Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 11: 149-160.
24. Morris AP, Lindgren CM, Zeggini E, Timpson NJ, Frayling TM, et al. (2010) A powerful approach to sub-phenotype analysis in population-based genetic association studies. Genet Epidemiol 34: 335-343.
25. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, et al. (2012) Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 44: 269-276.
26. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, et al. (2011) Human metabolic individuality in biomedical and pharmaceutical research. Nature 477: 54-60.
27. Probst-Hensch NM (2010) Chronic age-related diseases share risk factors: do they share pathophysiological mechanisms and why does that matter? Swiss Med Wkly 140: w13072.
28. Inouye M, Ripatti S, Kettunen J, Lyytikainen LP, Oksala N, et al. (2012) Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet 8: e1002907.
29. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44: 369-375, S361-363.
30. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107.
31. Lappalainen T, Montgomery SB, Nica AC, Dermitzakis ET (2011) Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet 89: 459-463.
32. Khan Z, Bloom JS, Amini S, Singh M, Perlman DH, et al. (2012) Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. Mol Syst Biol 8: 602.
33. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, et al. (2012) Genomics: ENCODE explained. Nature 489: 52-55.
34. Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong Purifying Selection at Synonymous Sites in D. melanogaster. PLoS Genet 9: e1003527.
35. Fridley BL, Biernacka JM (2011) Gene set analysis of SNP data: benefits, challenges, and future directions. Eur J Hum Genet 19: 837-843.
36. Petersen A, Alvarez C, Declaire S, Tintle NL (2013) Assessing Methods for Assigning SNPs to Genes in Gene-Based Tests of Association Using Common Variants. PLoS One 8: e62161.
37. Basu S, Pan W (2011) Comparison of statistical tests for disease association with rare variants. Genet Epidemiol 35: 606-619.
38. Liu DJ, Leal SM (2012) Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. Am J Hum Genet 91: 585-596.
39. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, et al. (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet 9: e1003520.
40. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 109: 1193-1198.
41. Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L (2013) Finding the sources of missing heritability in a yeast cross. Nature 494: 234-237.

42. Bhattacharya K, McCarthy MI, Morris AP (2011) Rapid testing of gene-gene interactions in genome-wide association studies of binary and quantitative phenotypes. Genet Epidemiol 35: 800-808.

43. Valdar W, Solberg LC, Gauguier D, Cookson WO, Rawlins JN, et al. (2006) Genetic and environmental effects on complex traits in mice. Genetics 174: 959-984.

44. Kauffmann F, Nadif R (2010) Candidate gene-environment interactions. J Epidemiol Community Health 64: 188-189.

45. Thomas D (2010) Gene--environment-wide association studies: emerging approaches. Nat Rev Genet 11: 259-272.

46. Risch N, Herrell R, Lehner T, Liang KY, Eaves L, et al. (2009) Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. JAMA 301: 2462-2471.

47. Smith PG, Day NE (1984) The design of case-control studies: the influence of confounding and interaction effects. Int J Epidemiol 13: 356-365.

48. VanderWeele TJ (2009) Sufficient cause interactions and statistical interactions. Epidemiology 20: 6-13.

49. Greenland S (2009) Interactions in epidemiology: relevance, identification, and estimation. Epidemiology 20: 14-17.

50. Moore JH, Asselbergs FW, Williams SM (2010) Bioinformatics challenges for genome-wide association studies. Bioinformatics 26: 445-455.

51. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet 44: 483-489.

52. Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61: 437-455.

53. Furrow RE, Christiansen FB, Feldman MW (2011) Environment-sensitive epigenetics and the heritability of complex diseases. Genetics 189: 1377-1387.

54. Rodenhiser D, Mann M (2006) Epigenetics and human disease: translating basic biology into clinical applications. CMAJ 174: 341-348.

55. Davies DE, Wicks J, Powell RM, Puddicombe SM, Holgate ST (2003) Airway remodeling in asthma: new insights. J Allergy Clin Immunol 111: 215-225; quiz 226.

56. Lessard A, Turcotte H, Cormier Y, Boulet LP (2008) Obesity and asthma: a specific phenotype? Chest 134: 317-323.

57. Barnes PJ (2000) Mechanisms in COPD: differences from asthma. Chest 117: 10S-14S.

58. Fabbri LM, Romagnoli M, Corbetta L, Casoni G, Busljetic K, et al. (2003) Differences in airway inflammation in patients with fixed airflow obstruction due to asthma or chronic obstructive pulmonary disease. Am J Respir Crit Care Med 167: 418-424.

59. Guerra S (2009) Asthma and chronic obstructive pulmonary disease. Curr Opin Allergy Clin Immunol 9: 409-416.

60. de Marco R, Pesce G, Marcon A, Accordini S, Antonicelli L, et al. (2013) The coexistence of asthma and chronic obstructive pulmonary disease (COPD): prevalence and risk factors in young, middle-aged and elderly people from the general population. PLoS One 8: e62985.

61. Holz O, Jorres RA, Timm P, Mucke M, Richter K, et al. (1999) Ozone-induced airway inflammatory changes differ between individuals and are reproducible. Am J Respir Crit Care Med 159: 776-784.

62. Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD (1990) Genetics of asthma and hay fever in Australian twins. Am Rev Respir Dis 142: 1351-1358.

63. Ingebrigtsen T, Thomsen SF, Vestbo J, van der Sluis S, Kyvik KO, et al. (2010) Genetic influences on Chronic Obstructive Pulmonary Disease - a twin study. Respir Med 104: 1890-1895.

64. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma P (2004) The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy 59: 469-478.

65. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 3: e442.

66. Jenkins MA, Clarke JR, Carlin JB, Robertson CF, Hopper JL, et al. (1996) Validation of questionnaire and bronchial hyperresponsiveness against respiratory physician assessment in the diagnosis of asthma. Int J Epidemiol 25: 609-616.

67. Agusti A, Calverley PM, Celli B, Coxson HO, Edwards LD, et al. (2010) Characterisation of COPD heterogeneity in the ECLIPSE cohort. Respir Res 11: 122.

68. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, et al. (2013) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 187: 347-365.

69. de Marco R, Locatelli F, Sunyer J, Burney P (2000) Differences in incidence of reported asthma related to age in men and women. A retrospective analysis of the data of the European Respiratory Health Survey. Am J Respir Crit Care Med 162: 68-74.

70. Romieu I, Fabre A, Fournier A, Kauffmann F, Varraso R, et al. (2010) Postmenopausal hormone therapy and asthma onset in the E3N cohort. Thorax 65: 292-297.

71. Moorman JE, Zahran H, Truman BI, Molla MT, Centers for Disease C, et al. (2011) Current asthma prevalence - United States, 2006-2008. MMWR Surveill Summ 60 Suppl: 84-86.

72. Eder W, Ege MJ, von Mutius E (2006) Current concepts: The asthma epidemic. N Engl J Med 355: 2226-2235.

73. Beuther DA, Sutherland ER (2007) Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. Am J Respir Crit Care Med 175: 661-666.

74. Neuman A, Hohmann C, Orsini N, Pershagen G, Eller E, et al. (2012) Maternal smoking in pregnancy and asthma in preschool children: a pooled analysis of eight birth cohorts. Am J Respir Crit Care Med 186: 1037-1043.

75. Ober C, Hoffjan S (2006) Asthma genetics 2006: the long and winding road to gene discovery. Genes Immun 7: 95-100.

76. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 448: 470-473.

77. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, et al. (2008) Effect of 17q21 variants and smoking exposure in early-onset asthma. N Engl J Med 359: 1985-1994.

78. Wu H, Romieu I, Shi M, Hancock DB, Li H, et al. (2010) Evaluation of candidate genes in a genome-wide association study of childhood asthma in Mexicans. J Allergy Clin Immunol 125: 321-327 e313.

79. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, et al. (2010) A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363: 1211-1221.

80. Ferreira MA, Matheson MC, Duffy DL, Marks GB, Hui J, et al. (2011) Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet 378: 1006-1014.

81. 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.

82. Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, et al. (2010) A genome-wide association study on African-ancestry populations for asthma. J Allergy Clin Immunol 125: 336-346 e334.

83. Kantor DB, Palmer CD, Young TR, Meng Y, Gajdos ZK, et al. (2013) Replication and fine mapping of asthma-associated loci in individuals of African ancestry. Hum Genet 132: 1039-1047.

84. Siroux V, Basagana X, Boudier A, Pin I, Garcia-Aymerich J, et al. (2011) Identifying adult asthma phenotypes using a clustering approach. Eur Respir J 38: 310-317.

85. Torgerson DG, Capurso D, Mathias RA, Graves PE, Hernandez RD, et al. (2012) Resequencing candidate genes implicates rare variants in asthma susceptibility. Am J Hum Genet 90: 273-281.

86. 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.

87. Li L, Kabesch M, Bouzigon E, Demenais F, Farrall M, et al. (2013) Using eQTL weights to improve power for genome-wide association studies: a genetic study of childhood asthma. Front Genet 4: 103.

88. Cameron L, Webster RB, Strempel JM, Kiesler P, Kabesch M, et al. (2006) Th2 cell-selective enhancement of human IL13 transcription by IL13-1112C>T, a polymorphism associated with allergic inflammation. J Immunol 177: 8633-8642.

89. Prefontaine D, Lajoie-Kadoch S, Foley S, Audusseau S, Olivenstein R, et al. (2009) Increased expression of IL-33 in severe asthma: evidence of expression by airway smooth muscle cells. J Immunol 183: 5094-5103.

90. Gerasimova A, Chavez L, Li B, Seumois G, Greenbaum J, et al. (2013) Predicting cell types and genetic variations contributing to disease by combining GWAS and epigenetic data. PLoS One 8: e54359.

91. Song GG, Lee YH (2013) Pathway analysis of genome-wide association study on asthma. Hum Immunol 74: 256-260.

92. von Mutius E (2009) Gene-environment interactions in asthma. J Allergy Clin Immunol 123: 3-11; quiz 12-13.

93. Jones MG (2008) Exposure-response in occupational allergy. Curr Opin Allergy Clin Immunol 8: 110-114.

94. 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.

95. Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, et al. (2012) Genome-wide association studies of asthma in population-based cohorts confirm known and suggested loci and identify an additional association near HLA. PLoS One 7: e44008.

96. Celli BR, MacNee W, Force AET (2004) Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 23: 932-946.

97. Salvi SS, Barnes PJ (2009) Chronic obstructive pulmonary disease in non-smokers. Lancet 374: 733-743.

98. Schikowski T, Mills IC, Anderson HR, Cohen A, Hansell A, et al. (2013) Ambient air pollution- a cause for COPD? Eur Respir J. Published Online First: 07 March 2013.

99. Mehta AJ, Miedinger D, Keidel D, Bettschart R, Bircher A, et al. (2012) Occupational exposure to dusts, gases, and fumes and incidence of chronic obstructive pulmonary disease in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults. Am J Respir Crit Care Med 185: 1292-1300.

100. Dahl M, Vestbo J, Zacho J, Lange P, Tybjaerg-Hansen A, et al. (2011) C reactive protein and chronic obstructive pulmonary disease: a Mendelian randomisation approach. Thorax 66: 197-204.

101. Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, et al. (2005) Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. Am J Respir Cell Mol Biol 33: 71-78.

102. Castaldi PJ, Cho MH, Cohn M, Langerman F, Moran S, et al. (2010) The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. Hum Mol Genet 19: 526-534.

103. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, et al. (2002) Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet 70: 1229-1239.

104. Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, et al. (2002) Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Hum Mol Genet 11: 623-632.

105. Zhu G, Warren L, Aponte J, Gulsvik A, Bakke P, et al. (2007) The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. Am J Respir Crit Care Med 176: 167-173.

106. Hersh CP, Silverman EK, Gascon J, Bhattacharya S, Klanderman BJ, et al. (2011) SOX5 is a candidate gene for chronic obstructive pulmonary disease susceptibility and is necessary for lung development. Am J Respir Crit Care Med 183: 1482-1489.

107. 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.

108. 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.

109. 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.

110. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, et al. (2010) Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 42: 200-202.

111. Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, et al. (2010) Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 182: 1498-1505.

112. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, et al. (2011) Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. Am J Respir Crit Care Med 183: 43-49.

113. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, et al. (2012) A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet 21: 947-957.

114. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, et al. (2012) Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. Am J Respir Crit Care Med 186: 622-632.

115. Castaldi PJ, Cho MH, Litonjua AA, Bakke P, Gulsvik A, et al. (2011) The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. Am J Respir Cell Mol Biol 45: 1147-1153.

116. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, et al. (2012) Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Hum Mol Genet 21: 1325-1335.

117. Qiu W, Baccarelli A, Carey VJ, Boutaoui N, Bacherman H, et al. (2012) Variable DNA methylation is associated with chronic obstructive pulmonary disease and lung function. Am J Respir Crit Care Med 185: 373-381.

118. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, et al. (2009) MMP12, lung function, and COPD in high-risk populations. N Engl J Med 361: 2599-2608.

119. Minelli C, Wei I, Sagoo G, Jarvis D, Shaheen S, et al. (2011) Interactive effects of antioxidant genes and air pollution on respiratory function and airway disease: a HuGE review. Am J Epidemiol 173: 603-620.

120. Engstrom G, Lind P, Hedblad B, Wollmer P, Stavenow L, et al. (2002) Lung function and cardiovascular risk: relationship with inflammation-sensitive plasma proteins. Circulation 106: 2555-2560.

121. Young RP, Hopkins R, Eaton TE (2007) Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. Eur Respir J 30: 616-622.

122. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al. (2005) Standardisation of spirometry. Eur Respir J 26: 319-338.

123. Cosio M, Ghezzo H, Hogg JC, Corbin R, Loveland M, et al. (1978) The relations between structural changes in small airways and pulmonary-function tests. N Engl J Med 298: 1277-1281.
124. Chen Y (1999) Genetics and pulmonary medicine.10: Genetic epidemiology of pulmonary function. Thorax 54: 818-824.
125. DeMeo DL, Carey VJ, Chapman HA, Reilly JJ, Ginns LC, et al. (2004) Familial aggregation of FEF(25-75) and FEF(25-75)/FVC in families with severe, early onset COPD. Thorax 59: 396-400.
126. Gottlieb DJ, Wilk JB, Harmon M, Evans JC, Joost O, et al. (2001) Heritability of longitudinal change in lung function. The Framingham study. Am J Respir Crit Care Med 164: 1655-1659.
127. Wilk JB, DeStefano AL, Joost O, Myers RH, Cupples LA, et al. (2003) Linkage and association with pulmonary function measures on chromosome 6q27 in the Framingham Heart Study. Hum Mol Genet 12: 2745-2751.
128. Wilk JB, Herbert A, Shoemaker CM, Gottlieb DJ, Karamohamed S (2007) Secreted modular calcium-binding protein 2 haplotypes are associated with pulmonary function. Am J Respir Crit Care Med 175: 554-560.
129. 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.
130. Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, et al. (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS One 6: e19382.
131. 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.
132. 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: e1003098.
133. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, et al. (2012) Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol 129: 1218-1228.
134. Hansel NN, Ruczinski I, Rafaels N, Sin DD, Daley D, et al. (2013) Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Hum Genet 132: 79-90.
135. Laurell CB, Eriksson A (1963) The electrophoretic alpha-1 globulin pattern of serum in alpha-1 antitrypsin deficiency. Scand J Clin Lab Invest 15: 132-140.
136. Senn O, Russi EW, Schindler C, Imboden M, von Eckardstein A, et al. (2008) Circulating alpha1-antitrypsin in the general population: determinants and association with lung function. Respir Res 9: 35.
137. Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, et al. (2001) Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. Atherosclerosis 154: 681-689.
138. Curiel DT, Chytil A, Courtney M, Crystal RG (1989) Serum alpha 1-antitrypsin deficiency associated with the common S-type (Glu264----Val) mutation results from intracellular degradation of alpha 1-antitrypsin prior to secretion. J Biol Chem 264: 10477-10486.
139. Lomas DA, Evans DL, Finch JT, Carrell RW (1992) The mechanism of Z alpha 1-antitrypsin accumulation in the liver. Nature 357: 605-607.
140. Sveger T (1976) Liver disease in alpha1-antitrypsin deficiency detected by screening of 200,000 infants. N Engl J Med 294: 1316-1321.
141. Eriksson S, Carlson J, Velez R (1986) Risk of cirrhosis and primary liver cancer in alpha 1-antitrypsin deficiency. N Engl J Med 314: 736-739.

142. Brantly ML, Wittes JT, Vogelmeier CF, Hubbard RC, Fells GA, et al. (1991) Use of a highly purified alpha 1-antitrypsin standard to establish ranges for the common normal and deficient alpha 1-antitrypsin phenotypes. Chest 100: 703-708.

143. Luisetti M, Seersholm N (2004) Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax 59: 164-169.

144. Dickens JA, Lomas DA (2011) Why has it been so difficult to prove the efficacy of alpha-1-antitrypsin replacement therapy? Insights from the study of disease pathogenesis. Drug Des Devel Ther 5: 391-405.

145. Tanash HA, Nilsson PM, Nilsson JA, Piitulainen E (2008) Clinical course and prognosis of never-smokers with severe alpha-1-antitrypsin deficiency (PiZZ). Thorax 63: 1091-1095.

146. Piitulainen E, Tornling G, Eriksson S (1997) Effect of age and occupational exposure to airway irritants on lung function in non-smoking individuals with alpha 1-antitrypsin deficiency (PiZZ). Thorax 52: 244-248.

147. Demeo DL, Sandhaus RA, Barker AF, Brantly ML, Eden E, et al. (2007) Determinants of airflow obstruction in severe alpha-1-antitrypsin deficiency. Thorax 62: 806-813.

148. Wood AM, Harrison RM, Semple S, Ayres JG, Stockley RA (2009) Outdoor air pollution is associated with disease severity in alpha1-antitrypsin deficiency. Eur Respir J 34: 346-353.

149. Wood AM, Harrison RM, Semple S, Ayres JG, Stockley RA (2010) Outdoor air pollution is associated with rapid decline of lung function in alpha-1-antitrypsin deficiency. Occup Environ Med 67: 556-561.

150. DeMeo DL, Campbell EJ, Brantly ML, Barker AF, Eden E, et al. (2009) Heritability of lung function in severe alpha-1 antitrypsin deficiency. Hum Hered 67: 38-45.

151. Dahl M, Hersh CP, Ly NP, Berkey CS, Silverman EK, et al. (2005) The protease inhibitor PI*S allele and COPD: a meta-analysis. Eur Respir J 26: 67-76.

152. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, et al. (2004) Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax 59: 843-849.

153. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, et al. (2010) alpha(1)-Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. Chest 138: 1125-1132.

154. Silva GE, Sherrill DL, Guerra S, Barbee RA (2003) A longitudinal study of alpha1-antitrypsin phenotypes and decline in FEV1 in a community population. Chest 123: 1435-1440.

155. Wadsworth ME, Vinall LE, Jones AL, Hardy RJ, Whitehouse DB, et al. (2004) Alpha1-antitrypsin as a risk for infant and adult respiratory outcomes in a national birth cohort. Am J Respir Cell Mol Biol 31: 559-564.

156. Sandford AJ, Chagani T, Weir TD, Connett JE, Anthonisen NR, et al. (2001) Susceptibility genes for rapid decline of lung function in the lung health study. Am J Respir Crit Care Med 163: 469-473.

157. Seersholm N, Wilcke JT, Kok-Jensen A, Dirksen A (2000) Risk of hospital admission for obstructive pulmonary disease in alpha(1)-antitrypsin heterozygotes of phenotype PiMZ. Am J Respir Crit Care Med 161: 81-84.

158. Dahl M, Tybjaerg-Hansen A, Lange P, Vestbo J, Nordestgaard BG (2002) Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha1-antitrypsin MZ heterozygotes: A longitudinal study of the general population. Ann Intern Med 136: 270-279.

159. Chapman KR, Stockley RA, Dawkins C, Wilkes MM, Navickis RJ (2009) Augmentation therapy for alpha1 antitrypsin deficiency: a meta-analysis. COPD 6: 177-184.

160. Gotzsche PC, Johansen HK (2010) Intravenous alpha-1 antitrypsin augmentation therapy for treating patients with alpha-1 antitrypsin deficiency and lung disease. Cochrane Database Syst Rev: CD007851.

161. Stockley RA, Parr DG, Piitulainen E, Stolk J, Stoel BC, et al. (2010) Therapeutic efficacy of alpha-1 antitrypsin augmentation therapy on the loss of lung tissue: an integrated analysis of 2 randomised clinical trials using computed tomography densitometry. Respir Res 11: 136.

162. Mulgrew AT, Taggart CC, Lawless MW, Greene CM, Brantly ML, et al. (2004) Z alpha1-antitrypsin polymerizes in the lung and acts as a neutrophil chemoattractant. Chest 125: 1952-1957.

163. Mahadeva R, Atkinson C, Li Z, Stewart S, Janciauskiene S, et al. (2005) Polymers of Z alpha1-antitrypsin co-localize with neutrophils in emphysematous alveoli and are chemotactic in vivo. Am J Pathol 166: 377-386.

164. Janus ED, Phillips NT, Carrell RW (1985) Smoking, lung function, and alpha 1-antitrypsin deficiency. Lancet 1: 152-154.

165. Piitulainen E, Eriksson S (1999) Decline in FEV1 related to smoking status in individuals with severe alpha1-antitrypsin deficiency (PiZZ). Eur Respir J 13: 247-251.

166. Silverman EK, Province MA, Campbell EJ, Pierce JA, Rao DC (1992) Family study of alpha 1-antitrypsin deficiency: effects of cigarette smoking, measured genotype, and their interaction on pulmonary function and biochemical traits. Genet Epidemiol 9: 317-331.

167. Taggart C, Cervantes-Laurean D, Kim G, McElvaney NG, Wehr N, et al. (2000) Oxidation of either methionine 351 or methionine 358 in alpha 1-antitrypsin causes loss of anti-neutrophil elastase activity. J Biol Chem 275: 27258-27265.

168. Alam S, Li Z, Janciauskiene S, Mahadeva R (2011) Oxidation of Z alpha1-antitrypsin by cigarette smoke induces polymerization: a novel mechanism of early-onset emphysema. Am J Respir Cell Mol Biol 45: 261-269.

169. Senn O, Russi EW, Imboden M, Probst-Hensch NM (2005) alpha1-Antitrypsin deficiency and lung disease: risk modification by occupational and environmental inhalants. Eur Respir J 26: 909-917.

170. Pabalan N, Bapat B, Sung L, Jarjanazi H, Francisco-Pabalan O, et al. (2008) Cyclin D1 Pro241Pro (CCND1-G870A) polymorphism is associated with increased cancer risk in human populations: a meta-analysis. Cancer Epidemiol Biomarkers Prev 17: 2773-2781.

171. Whibley C, Pharoah PD, Hollstein M (2009) p53 polymorphisms: cancer implications. Nat Rev Cancer 9: 95-107.

172. Imboden M, Schwartz J, Schindler C, Curjuric I, Berger W, et al. (2009) Decreased PM10 exposure attenuates age-related lung function decline: genetic variants in p53, p21, and CCND1 modify this effect. Environ Health Perspect 117: 1420-1427.

173. American Thoracic S, European Respiratory S (2003) American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. Am J Respir Crit Care Med 168: 818-900.

174. Dahl M, Nordestgaard BG, Lange P, Vestbo J, Tybjaerg-Hansen A (2001) Molecular diagnosis of intermediate and severe alpha(1)-antitrypsin deficiency: MZ individuals with chronic obstructive pulmonary disease may have lower lung function than MM individuals. Clin Chem 47: 56-62.

175. Silverman EK, Province MA, Rao DC, Pierce JA, Campbell EJ (1990) A family study of the variability of pulmonary function in alpha 1-antitrypsin deficiency. Quantitative phenotypes. Am Rev Respir Dis 142: 1015-1021.

176. Larsson C, Eriksson S, Dirksen H (1977) Smoking and intermediate alpha1-antitrypsin deficiency and lung function in middle-aged men. Br Med J 2: 922-925.

177. Pierre F, Pham QT, Mur JM, Chau N, Martin JP (1988) Respiratory symptoms and pulmonary function in 871 miners according to Pi phenotype: a longitudinal study. Eur J Epidemiol 4: 39-44.

178. Banauch GI, Brantly M, Izbicki G, Hall C, Shanske A, et al. (2010) Accelerated spirometric decline in New York City firefighters with alpha(1)-antitrypsin deficiency. Chest 138: 1116-1124.

179. Downs SH, Schindler C, Liu LJ, Keidel D, Bayer-Oglesby L, et al. (2007) Reduced exposure to PM10 and attenuated age-related decline in lung function. N Engl J Med 357: 2338-2347.

180. 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 Praventivmed 42: 67-84.

181. Ackermann-Liebrich U, Kuna-Dibbert B, Probst-Hensch NM, Schindler C, Felber Dietrich D, et al. (2005) Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. Soz Praventivmed 50: 245-263.

182. Kunzli N, Kuna-Dibbert B, Keidel D, Keller R, Brandli O, et al. (2005) Longitudinal validity of spirometers--a challenge in longitudinal studies. Swiss Med Wkly 135: 503-508.

183. Liu LJ, Curjuric I, Keidel D, Heldstab J, Kunzli N, et al. (2007) Characterization of source-specific air pollution exposure for a large population-based Swiss cohort (SAPALDIA). Environ Health Perspect 115: 1638-1645.

184. Probst-Hensch NM, Curjuric I, Pierre-Olivier B, Ackermann-Liebrich U, Bettschart RW, et al. (2010) Longitudinal change of prebronchodilator spirometric obstruction and health outcomes: results from the SAPALDIA cohort. Thorax 65: 150-156.

185. Byth BC, Billingsley GD, Cox DW (1994) Physical and genetic mapping of the serpin gene cluster at 14q32.1: allelic association and a unique haplotype associated with alpha 1-antitrypsin deficiency. Am J Hum Genet 55: 126-133.

186. Cox DW, Woo SL, Mansfield T (1985) DNA restriction fragments associated with alpha 1-antitrypsin indicate a single origin for deficiency allele PI Z. Nature 316: 79-81.

187. Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145.

188. Bobadilla JL, Macek M, Jr., Fine JP, Farrell PM (2002) Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat 19: 575-606.

189. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, et al. (2008) Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 40: 592-599.

190. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324: 387-389.

191. Zhu Q, Ge D, Maia JM, Zhu M, Petrovski S, et al. (2011) A genome-wide comparison of the functional properties of rare and common genetic variants in humans. Am J Hum Genet 88: 458-468.

192. McClellan J, King MC (2010) Genetic heterogeneity in human disease. Cell 141: 210-217.

193. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. PLoS Biol 8: e1000294.

194. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, et al. (2013) All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLoS Genet 9: e1003449.

195. Lomas DA (2006) The selective advantage of alpha1-antitrypsin deficiency. Am J Respir Crit Care Med 173: 1072-1077.

196. Anderson CA, Soranzo N, Zeggini E, Barrett JC (2011) Synthetic associations are unlikely to account for many common disease genome-wide association signals. PLoS Biol 9: e1000580.

197. Wray NR, Purcell SM, Visscher PM (2011) Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol 9: e1000579.

198. Marigorta UM, Navarro A (2013) High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. PLoS Genet 9: e1003566.

199. Ntzani EE, Liberopoulos G, Manolio TA, Ioannidis JP (2012) Consistency of genome-wide associations across major ancestral groups. Hum Genet 131: 1057-1071.

200. Sanna S, Li B, Mulas A, Sidore C, Kang HM, et al. (2011) Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. PLoS Genet 7: e1002198.

201. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, et al. (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43: 1066-1073.

202. Ramsey LB, Bruun GH, Yang W, Trevino LR, Vattathil S, et al. (2012) Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. Genome Res 22: 1-8.

203. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, et al. (2013) Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 498: 232-235.

204. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64-69.

205. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, et al. (2012) An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337: 100-104.

206. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, et al. (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493: 216-220.

207. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569.

208. Vinkhuyzen AA, Wray NR, Yang J, Goddard ME, Visscher PM (2013) Estimation and Partition of Heritability in Human Populations Using Whole-Genome Analysis Methods. Annu Rev Genet. Published Online First: 22 August 2013.

209. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, et al. (2013) Heritability of COPD and Related Phenotypes in Smokers. Am J Respir Crit Care Med 188: 941-947.

210. Curjuric I, Imboden M, Schindler C, Downs SH, Hersberger M, et al. (2010) HMOX1 and GST variants modify attenuation of FEF25-75% decline due to PM10 reduction. Eur Respir J 35: 505-514.

211. Donato LJ, Jenkins SM, Smith C, Katzmann JA, Snyder MR (2012) Reference and interpretive ranges for alpha(1)-antitrypsin quantitation by phenotype in adult and pediatric populations. Am J Clin Pathol 138: 398-405.

212. Bornhorst JA, Greene DN, Ashwood ER, Grenache DG (2013) alpha1-Antitrypsin phenotypes and associated serum protein concentrations in a large clinical population. Chest 143: 1000-1008.

213. Ferrarotti I, Thun GA, Probst-Hensch NM, Luisetti M (2013) alpha1-Antitrypsin Level and Pheno/Genotypes. Chest 144: 1732-1733.

214. Nickels S, Truong T, Hein R, Stevens K, Buck K, et al. (2013) Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet 9: e1003284.

215. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, et al. (2009) Rapid DNA methylation changes after exposure to traffic particles. Am J Respir Crit Care Med 179: 572-578.

216. Shore SA, Johnston RA (2006) Obesity and asthma. Pharmacol Ther 110: 83-102.

217. Medina-Tato DA, Ward SG, Watson ML (2007) Phosphoinositide 3-kinase signalling in lung disease: leucocytes and beyond. Immunology 121: 448-461.

218. Melen E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, et al. (2013) Genome-wide association study of body mass index in 23 000 individuals with and without asthma. Clin Exp Allergy 43: 463-474.

219. Chang CQ, Yesupriya A, Rowell JL, Pimentel CB, Clyne M, et al. (2013) A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. Eur J Hum Genet. Published Online First: 24 July 2013.

220. Vimaleswaran KS, Tachmazidou I, Zhao JH, Hirschhorn JN, Dudbridge F, et al. (2012) Candidate genes for obesity-susceptibility show enriched association within a large genome-wide association study for BMI. Hum Mol Genet 21: 4537-4542.

221. Collins AL, Kim Y, Sklar P, International Schizophrenia C, O'Donovan MC, et al. (2012) Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. Psychol Med 42: 607-616.

222. Li L, Li Y, Browning SR, Browning BL, Slater AJ, et al. (2011) Performance of genotype imputation for rare variants identified in exons and flanking regions of genes. PLoS One 6: e24945.

223. Yao TC, Du G, Han L, Sun Y, Hu D, et al. (2013) Genome-wide association study of lung function phenotypes in a founder population. J Allergy Clin Immunol. Published Online First: 06 August 2013.

224. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485: 237-241.

225. Cohorts for H, Aging Research in Genetic Epidemiology C, Morrison AC, Voorman A, Johnson AD, et al. (2013) Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. Nat Genet 45: 899-901.

226. Wetterstrand KA DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcosts. Accessed: 01 September 2013.

227. Lee S, Teslovich TM, Boehnke M, Lin X (2013) General Framework for Meta-analysis of Rare Variants in Sequencing Association Studies. Am J Hum Genet 93: 42-53.

228. Fellay J, Thompson AJ, Ge D, Gumbs CE, Urban TJ, et al. (2010) ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. Nature 464: 405-408.

229. Group SC, Link E, Parish S, Armitage J, Bowman L, et al. (2008) SLCO1B1 variants and statin-induced myopathy--a genomewide study. N Engl J Med 359: 789-799.

230. Baldwin RM, Owzar K, Zembutsu H, Chhibber A, Kubo M, et al. (2012) A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. Clin Cancer Res 18: 5099-5109.

231. Cooper GM, Johnson JA, Langaee TY, Feng H, Stanaway IB, et al. (2008) A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. Blood 112: 1022-1027.

232. Yang JJ, Cheng C, Yang WJ, Pei DQ, Cao XY, et al. (2009) Genome-wide Interrogation of Germline Genetic Variation Associated With Treatment Response in Childhood Acute Lymphoblastic Leukemia. JAMA 301: 393-403.

233. Mallal S, Phillips E, Carosi G, Molina JM, Workman C, et al. (2008) HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med 358: 568-579.

234. Yan J, Aliev F, Webb BT, Kendler KS, Williamson VS, et al. (2013) Using genetic information from candidate gene and genome-wide association studies in risk prediction for alcohol dependence. Addict Biol. Published Online First: 30 January 2013.

235. Spycher BD, Henderson J, Granell R, Evans DM, Smith GD, et al. (2012) Genome-wide prediction of childhood asthma and related phenotypes in a longitudinal birth cohort. J Allergy Clin Immunol 130: 503-509 e507.

236. Jostins L, Barrett JC (2011) Genetic risk prediction in complex disease. Hum Mol Genet 20: R182-188.

237. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579-589.

238. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, et al. (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 340: b4838.

239. Do CB, Hinds DA, Francke U, Eriksson N (2012) Comparison of family history and SNPs for predicting risk of complex disease. PLoS Genet 8: e1002973.

240. Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. PLoS Genet 9: e1003348.

241. Chatenoud L, Warncke K, Ziegler AG (2012) Clinical immunologic interventions for the treatment of type 1 diabetes. Cold Spring Harb Perspect Med 2.

242. Patel CJ, Sivadas A, Tabassum R, Preeprem T, Zhao J, et al. (2013) Whole Genome Sequencing in support of Wellness and Health Maintenance. Genome Med 5: 58.

243. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, et al. (2010) Clinical assessment incorporating a personal genome. Lancet 375: 1525-1535.

244. Bainbridge MN, Wiszniewski W, Murdock DR, Friedman J, Gonzaga-Jauregui C, et al. (2011) Whole-genome sequencing for optimized patient management. Sci Transl Med 3: 87re83.

245. van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, et al. (2013) Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. Eur J Hum Genet 21: 580-584.

246. Hayden EC (2013) Geneticists push for global data-sharing. Nature 498: 16-17.

247. Fears R, Ter Meulen V (2013) The perspective from EASAC and FEAM on direct-to-consumer genetic testing for health-related purposes. Eur J Hum Genet 21: 703-707.

248. Borry P, van Hellemondt RE, Sprumont D, Jales CFD, Rial-Sebbag E, et al. (2012) Legislation on direct-to-consumer genetic testing in seven European countries. Eur J Hum Genet 20: 715-721.

249. Tung JY, Do CB, Hinds DA, Kiefer AK, Macpherson JM, et al. (2011) Efficient replication of over 180 genetic associations with self-reported medical data. PLoS One 6: e23473.

250. Janssens AC, Kraft P (2012) Research conducted using data obtained through online communities: ethical implications of methodological limitations. PLoS Med 9: e1001328.

251. Howard HC, Borry P (2013) Survey of European clinical geneticists on awareness, experiences and attitudes towards direct-to-consumer genetic testing. Genome Med 5: 45.

252. Bloss CS, Darst BF, Topol EJ, Schork NJ (2011) Direct-to-consumer personalized genomic testing. Hum Mol Genet 20: R132-141.

253. Christensen KD, Jayaratne TE, Roberts JS, Kardia SL, Petty EM (2010) Understandings of basic genetics in the United States: results from a national survey of black and white men and women. Public Health Genomics 13: 467-476.

254. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, et al. (2011) Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet 20: 4786-4796.

255. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 82: 763-771.

256. Forsberg LA, Absher D, Dumanski JP (2013) Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. J Med Genet 50: 1-10.

257. Zimmern RL, Khoury MJ (2012) The Impact of Genomics on Public Health Practice: The Case for Change. Public Health Genomics 15: 118-124.

258. Knowles JW, Assimes TL, Kiernan M, Pavlovic A, Goldstein BA, et al. (2012) Randomized trial of personal genomics for preventive cardiology: design and challenges. Circ Cardiovasc Genet 5: 368-376.

259. Pencina MJ, D'Agostino RB, Sr., Larson MG, Massaro JM, Vasan RS (2009) Predicting the 30-year risk of cardiovascular disease: the framingham heart study. Circulation 119: 3078-3084.

260. Patel CJ, Bhattacharya J, Butte AJ (2010) An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS One 5: e10746.

261. Watts C, Cairncross S (2012) Should the GBD risk factor rankings be used to guide policy? Lancet 380: 2060-2061.

262. Alwan A, MacLean DR (2009) A review of non-communicable disease in low- and middle-income countries. Int Health 1: 3-9.

263. Stuckler D, McKee M, Ebrahim S, Basu S (2012) Manufacturing Epidemics: The Role of Global Producers in Increased Consumption of Unhealthy Commodities Including Processed Foods, Alcohol, and Tobacco. PloS Med 9.

264. Beaglehole R, Ebrahim S, Reddy S, Voute J, Leeder S, et al. (2007) Chronic Diseases 5 - Prevention of chronic diseases: a call to action. Lancet 370: 2152-2157.

265. Chan M (2012) From new estimates to better data. Lancet 380: 2054.

266. 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.

267. 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.

268. 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.

269. Cornelis MC, 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.

270. Clayton D, McKeigue PM (2001) Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet 358: 1356-1360.

271. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148: 1293-1307.

272. Karmaus W, Ziyab AH, Everson T, Holloway JW (2013) Epigenetic mechanisms and models in the origins of asthma. Curr Opin Allergy Clin Immunol 13: 63-69.

273. Mayer AS, Stoller JK, Bucher Bartelson B, James Ruttenber A, Sandhaus RA, et al. (2000) Occupational exposure risks in individuals with PI*Z alpha(1)-antitrypsin deficiency. Am J Respir Crit Care Med 162: 553-558.

274. Blanco I, de Serres FJ, Carcaba V, Lara B, Fernandez-Bustillo E (2012) Alpha-1 Antitrypsin Deficiency PI*Z and PI*S Gene Frequency Distribution Using on Maps of the World by an Inverse Distance Weighting (IDW) Multivariate Interpolation Method. Hepat Mon 12: e7434.

275. Greulich T, Ottaviani S, Bals R, Lepper PM, Vogelmeier C, et al. (2013) Alpha1-antitrypsin deficiency - Diagnostic testing and disease awareness in Germany and Italy. Respir Med 107: 1400-1408.

276. Stoller JK, Brantly M (2013) The challenge of detecting alpha-1 antitrypsin deficiency. COPD 10 Suppl 1: 26-34.

277. Stoller JK, Sandhaus RA, Turino G, Dickson R, Rodgers K, et al. (2005) Delay in diagnosis of alpha1-antitrypsin deficiency: a continuing problem. Chest 128: 1989-1994.

278. Francavilla R, Castellaneta SP, Hadzic N, Chambers SM, Portmann B, et al. (2000) Prognosis of alpha-1-antitrypsin deficiency-related liver disease in the era of paediatric liver transplantation. J Hepatol 32: 986-992.

279. Thelin T, Sveger T, McNeil TF (1996) Primary prevention in a high-risk group: smoking habits in adolescents with homozygous alpha-1-antitrypsin deficiency (ATD). Acta Paediatr 85: 1207-1212.

280. Sveger T, Thelin T, McNeil TF (1997) Young adults with alpha 1-antitrypsin deficiency identified neonatally: their health, knowledge about and adaptation to the high-risk condition. Acta Paediatr 86: 37-40.

281. Sveger T, Thelin T (2000) A future for neonatal alpha1-antitrypsin screening? Acta Paediatr 89: 628-631.

282. Wang D, Wang W, Dawkins P, Paterson T, Kalsheker N, et al. (2011) Deletion of Serpina1a, a murine alpha1-antitrypsin ortholog, results in embryonic lethality. Exp Lung Res 37: 291-300.

283. Mueller C, Flotte TR (2013) Gene-based therapy for alpha-1 antitrypsin deficiency. COPD 10 Suppl 1: 44-49.

284. Lewis EC (2012) Expanding the clinical indications for alpha(1)-antitrypsin therapy. Mol Med 18: 957-970.

285. Hidvegi T, Ewing M, Hale P, Dippold C, Beckett C, et al. (2010) An autophagy-enhancing drug promotes degradation of mutant alpha1-antitrypsin Z and reduces hepatic fibrosis. Science 329: 229-232.

286. Castaldi PJ, DeMeo DL, Kent DM, Campbell EJ, Barker AF, et al. (2009) Development of predictive models for airflow obstruction in alpha-1-antitrypsin deficiency. Am J Epidemiol 170: 1005-1013.

287. Kim WJ, Wood AM, Barker AF, Brantly ML, Campbell EJ, et al. (2012) Association of IREB2 and CHRNA3 polymorphisms with airflow obstruction in severe alpha-1 antitrypsin deficiency. Respir Res 13: 16.

288. Jonigk D, Al-Omari M, Maegel L, Muller M, Izykowski N, et al. (2013) Anti-inflammatory and immunomodulatory properties of alpha1-antitrypsin without inhibition of elastase. Proc Natl Acad Sci U S A 110: 15007-15012.

289. Talmud PJ, Martin S, Steiner G, Flavell DM, Whitehouse DB, et al. (2003) Progression of atherosclerosis is associated with variation in the alpha1-antitrypsin gene. Arterioscler Thromb Vasc Biol 23: 644-649.

290. Miller VM, Petterson TM, Jeavons EN, Lnu AS, Rider DN, et al. (2013) Genetic polymorphisms associated with carotid artery intima-media thickness and coronary artery calcification in women of the Kronos Early Estrogen Prevention Study. Physiol Genomics 45: 79-88.

291. Duckers JM, Shale DJ, Stockley RA, Gale NS, Evans BA, et al. (2010) Cardiovascular and musculskeletal co-morbidities in patients with alpha 1 antitrypsin deficiency. Respir Res 11: 173.

292. Vizzardi E, Corda L, Pezzali N, Roca E, Pini L, et al. (2012) Elastic properties of the ascending aorta in patients with alpha1-antitrypsin deficiency (Z homozygotes). Heart 98: 1354-1358.

293. Mashiba S, Wada Y, Takeya M, Sugiyama A, Hamakubo T, et al. (2001) In vivo complex formation of oxidized alpha(1)-antitrypsin and LDL. Arterioscler Thromb Vasc Biol 21: 1801-1808.

9 Appendix

CURRICULUM VITAE

PERSONAL DATA

Name	Gian Andri Thun
Date & Place of Birth	06/06/1976, Erlangen (Germany)
Nationality	Swiss

EDUCATION

10.2009 – 10.2013 **Swiss Tropical and Public Health Institute Basel, Doctoral Studies**
Thesis: Lung Function in the General Population: the Complex Interplay of Variants in *SERPINA1* and other Genes with the Environment

10.2003 – 01.2007 **University of Zürich, Teacher Education**
Diploma in Secondary and Higher Education in Biology

10.1996 – 06.2002 **University of Zürich, Studies in Biology**
Master in Molecular Biology (MSc), Minors: Biochemistry and Computer Science
Master Thesis: Prediction of Protein Interaction Sites from Sequence

08.1991 – 01.1996 **Gymnasium Zürich Oerlikon**
Matura, Type C (general qualification for university entrance)

WORK EXPERIENCE

08.2006 – 07.2009 University of Zürich, Institute of Social and Preventive Medicine
Technical Assistant (part-time)

08.2006 – 08.2009 Realgymnasium Rämibühl, Zürich
Teacher in Biology (part-time)

10.2005 – 07.2006 Kantonsschule Zug and Gymnasium Immensee
Substitute Teacher in Biology

TEACHING AND TRAINING

2013 Lecturer in Epidemiology, Module for Master in Advanced Studies in Food Safety, University of Basel

2013 Tutor in Scientific Writing and Presenting for Medical Students (Wissenschaftsmonat), University of Basel

2010 – 2013	Tutor in Statistics for Medical Students, University of Basel
2001 – 2004	Tutor in Biochemistry for Biology Students, University of Zürich
1998 – 1999	Tutor in Mathematics for Biology Students, University of Zürich

LIST OF PUBLICATIONS

Mehta AJ, **Thun GA**, Imboden M, Ferrarotti I, Keidel D, Künzli N, Kromhout H, Miedinger D, Phuleria H, Rochat T, Russi EW, Schindler C, Schwartz J, Vermeulen R, Luisetti M, Probst-Hensch N; SAPALDIA team. Interactions between *SERPINA1* PiMZ Genotype, Occupational Exposure and Lung Function Decline. *Occupational and Environmental Medicine*. Published Online First: 08 November 2013.

Ferrarotti I, **Thun GA**, Probst-Hensch NM, Luisetti M. Alpha1-Antitrypsin Level and Pheno/Genotypes. *Chest 2013*, 144(5), 1732-1733.

Parsa A, Fuchsberger C, Köttgen A, O'Seaghdha CM, Pattaro C, de Andrade M, Chasman DI, Teumer A, Endlich K, Olden M, Chen MH, Tin A, Kim YJ, Taliun D, Li M, Feitosa M, Gorski M, Yang Q, Hundertmark C, Foster MC, Glazer N, Isaacs A, Rao M, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson A, Tönjes A, Dehghan A, Couraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G, Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Hofer E, Hu F, Demirkan A, Oostra BA, Turner ST, Ding J, Andrews JS, Freedman BI, Giulianini F, Koenig W, Illig T, Döring A, Wichmann HE, Zgaga L, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Stengel B, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Colhoun H, Doney A, Robino A, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, **Thun GA**, Paulweber B, Haun M, Sala C, Mitchell P, Ciullo M, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman J, Hayward C, Ridker PM, Bochud M, Heid IM, Siscovick DS, Fox CS, Kao WL, Böger CA. Common Variants in Mendelian Kidney Disease Genes and their Association with Renal Function. *Journal of the American Society of Nephrology 2013*. Published Online First: 12 September 2013.

Thun GA, Imboden M, Ferrarotti I, Kumar A, Obeidat M, Zorzetto M, Haun M, Curjuric I, Couto Alves A, Jackson VE, Albrecht E, Ried JS, Teumer A, Lopez LM, Huffman JE, Enroth S, Bossé Y, Hao K, Timens W, Gyllensten U, Polasek O, Wilson JF, Rudan I, Hayward C, Sandford AJ, Deary IJ, Koch B, Reischl E, Schulz H, Hui J, James AL, Rochat T, Russi EW, Jarvelin MR, Strachan DP, Hall IP, Tobin MD, Dahl M, Fallgaard Nielsen S, Nordestgaard BG, Kronenberg F, Luisetti M, Probst-Hensch NM. Causal and Synthetic Associations of Variants in the *SERPINA* Gene Cluster with Alpha1-antitrypsin Serum Levels. *PLoS Genetics 2013*, 9(8), e1003585.

Thun GA, Imboden M, Berger W, Rochat T, Probst-Hensch NM. The Association of a Variant in the Cell Cycle Control Gene *CCND1* and Obesity on the Development of Asthma in the Swiss SAPALDIA Study. *Journal of Asthma 2013*, 50(2), 147-154.

Köttgen A, Albrecht E, Teumer A, Vitart V, Krumsiek J, Hundertmark C, Pistis G, Ruggiero D, O'Seaghdha CM, Haller T, Yang Q, Tanaka T, Johnson AD, Kutalik Z, Smith AV, Shi J, Struchalin M, Middelberg RP, Brown MJ, Gaffo AL, Pirastu N, Li G, Hayward C, Zemunik T, Huffman J, Yengo L, Zhao JH, Demirkan A, Feitosa MF, Liu X, Malerba G, Lopez LM, van der Harst P, Li X, Kleber ME, Hicks AA, Nolte IM, Johansson A, Murgia F, Wild SH, Bakker SJ, Peden JF, Dehghan A, Steri M, Tenesa A, Lagou V, Salo P, Mangino M, Rose LM, Lehtimäki T, Woodward OM, Okada Y, Tin A, Müller C, Oldmeadow C, Putku M, Czamara D, Kraft P, Frogheri L, **Thun GA**, Grotevendt A, Gislason GK, Harris TB, Launer LJ, McArdle P, Shuldiner AR, Boerwinkle E, Coresh J, Schmidt H, Schallert M, Martin NG, Montgomery GW, Kubo M, Nakamura Y, Tanaka T, Munroe PB, Samani NJ, Jacobs DR Jr, Liu K, D'Adamo P, Ulivi S, Rotter JI, Psaty BM, Vollenweider P, Waeber G, Campbell S, Devuyst O, Navarro P, Kolcic I, Hastie N, Balkau B, Froguel P, Esko T, Salumets A, Khaw KT, Langenberg C, Wareham NJ, Isaacs A, Kraja A, Zhang Q, Wild PS, Scott RJ, Holliday EG, Org E, Viigimaa M, Bandinelli S, Metter JE, Lupo A, Trabetti E, Sorice R, Döring A, Lattka E, Strauch K, Theis F, Waldenberger M, Wichmann HE, Davies G, Gow AJ, Bruinenberg M; LifeLines Cohort Study, Stolk RP, Kooner JS, Zhang W, Winkelmann BR, Boehm BO, Lucae S, Penninx BW, Smit JH, Curhan G, Mudgal P, Plenge RM, Portas L, Persico I, Kirin M, Wilson JF, Mateo Leach I, van Gilst WH, Goel A, Ongen H, Hofman A, Rivadeneira F, Uitterlinden AG, Imboden M, von Eckardstein A, Cucca F, Nagaraja R, Piras MG, Nauck M, Schurmann C, Budde K, Ernst F, Farrington SM, Theodoratou E, Prokopenko I, Stumvoll M, Jula A, Perola M, Salomaa V, Shin SY, Spector TD, Sala C, Ridker PM, Kähönen M, Viikari J, Hengstenberg C, Nelson CP; CARDIoGRAM Consortium; DIAGRAM Consortium; ICBP Consortium; MAGIC Consortium, Meschia JF, Nalls MA, Sharma P, Singleton AB, Kamatani N, Zeller T, Burnier M, Attia J, Laan M, Klopp N, Hillege HL, Kloiber S, Choi H, Pirastu M, Tore S, Probst-Hensch NM, Völzke H, Gudnason V, Parsa A, Schmidt R, Whitfield JB, Fornage M, Gasparini P, Siscovick DS, Polašek O, Campbell H, Rudan I, Bouatia-Naji N, Metspalu A, Loos RJ, van Duijn CM, Borecki IB, Ferrucci L, Gambaro G, Deary IJ, Wolffenbuttel BH, Chambers JC, März W, Pramstaller PP, Snieder H, Gyllensten U, Wright AF, Navis G, Watkins H, Witteman JC, Sanna S, Schipf S, Dunlop MG, Tönjes A, Ripatti S, Soranzo N, Toniolo D, Chasman DI, Raitakari O, Kao WH, Ciullo M, Fox CS, Caulfield M, Bochud M, Gieger C. Genome-wide Association Analyses Identify 18 New Loci Associated with Serum Urate Concentrations. *Nature Genetics 2013*, 45(2), 145-154.

Chasman DI, Fuchsberger C, Pattaro C, Teumer A, Böger CA, Endlich K, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa MF, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson A, Tönjes A, Dehghan A, Lambert JC, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Coassin S, Pistis G; CARDIoGRAM Consortium; ICBP Consortium; CARe Consortium; WTCCC2, Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu F, Demirkan A, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Giulianini F, Koenig W, Illig T, Meisinger C, Gieger C, Zgaga L, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Stengel B, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, **Thun GA**, Paulweber B, Haun M, Sala C, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki IB, Kardia SL, Liu Y,

Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman J, Hayward C, Ridker PM, Parsa A, Bochud M, Heid IM, Kao WH, Fox CS, Köttgen A. Integration of Genome-wide Association Studies with Biological Knowledge Identifies Six Novel Genes Related to Kidney Function. *Human Molecular Genetics 2012*, 21(24), 5329-5343.

Thun GA, Ferrarotti I, Imboden M, Rochat T, Gerbase M, Kronenberg F, Bridevaux PO, Zemp E, Zorzetto M, Ottaviani S, Russi EW, Luisetti M, Probst-Hensch NM. *SERPINA1* PiZ and PiS Heterozygotes and Lung Function Decline in the SAPALDIA Cohort. *PLoS ONE 2012*, 7(8), e42728.

Pattaro C, Köttgen A, Teumer A, Garnaas M, Böger CA, Fuchsberger C, Olden M, Chen MH, Tin A, Taliun D, Li M, Gao X, Gorski M, Yang Q, Hundertmark C, Foster MC, O'Seaghdha CM, Glazer N, Isaacs A, Liu CT, Smith AV, O'Connell JR, Struchalin M, Tanaka T, Li G, Johnson AD, Gierman HJ, Feitosa M, Hwang SJ, Atkinson EJ, Lohman K, Cornelis MC, Johansson Å, Tönjes A, Dehghan A, Chouraki V, Holliday EG, Sorice R, Kutalik Z, Lehtimäki T, Esko T, Deshmukh H, Ulivi S, Chu AY, Murgia F, Trompet S, Imboden M, Kollerits B, Pistis G; CARDIoGRAM Consortium; ICBP Consortium; CARe Consortium; Wellcome Trust Case Control Consortium 2 (WTCCC2), Harris TB, Launer LJ, Aspelund T, Eiriksdottir G, Mitchell BD, Boerwinkle E, Schmidt H, Cavalieri M, Rao M, Hu FB, Demirkan A, Oostra BA, de Andrade M, Turner ST, Ding J, Andrews JS, Freedman BI, Koenig W, Illig T, Döring A, Wichmann HE, Kolcic I, Zemunik T, Boban M, Minelli C, Wheeler HE, Igl W, Zaboli G, Wild SH, Wright AF, Campbell H, Ellinghaus D, Nöthlings U, Jacobs G, Biffar R, Endlich K, Ernst F, Homuth G, Kroemer HK, Nauck M, Stracke S, Völker U, Völzke H, Kovacs P, Stumvoll M, Mägi R, Hofman A, Uitterlinden AG, Rivadeneira F, Aulchenko YS, Polasek O, Hastie N, Vitart V, Helmer C, Wang JJ, Ruggiero D, Bergmann S, Kähönen M, Viikari J, Nikopensius T, Province M, Ketkar S, Colhoun H, Doney A, Robino A, Giulianini F, Krämer BK, Portas L, Ford I, Buckley BM, Adam M, **Thun GA**, Paulweber B, Haun M, Sala C, Metzger M, Mitchell P, Ciullo M, Kim SK, Vollenweider P, Raitakari O, Metspalu A, Palmer C, Gasparini P, Pirastu M, Jukema JW, Probst-Hensch NM, Kronenberg F, Toniolo D, Gudnason V, Shuldiner AR, Coresh J, Schmidt R, Ferrucci L, Siscovick DS, van Duijn CM, Borecki I, Kardia SL, Liu Y, Curhan GC, Rudan I, Gyllensten U, Wilson JF, Franke A, Pramstaller PP, Rettig R, Prokopenko I, Witteman JC, Hayward C, Ridker P, Parsa A, Bochud M, Heid IM, Goessling W, Chasman DI, Kao WH, Fox CS. Genome-wide Association and Functional Follow-up Reveals New Loci for Kidney Function. *PLoS Genetics 2012*, 8(3), e1002584.

Ferrarotti I, **Thun GA**, Zorzetto M, Ottaviani S, Imboden M, Schindler C, von Eckardstein A, Rohrer L, Rochat T, Russi EW, Probst-Hensch NM, Luisetti M. Serum Levels and Genotype Distribution of α1-antitrypsin in the General Population. *Thorax 2012*, 67(8), 669-674.

Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, Wilk JB, Vonk JM, **Thun GA**, Siroux V, Nadif R, Monier F, Gonzalez JR, Wjst M, Heinrich J, Loehr LR, Franceschini N, North KE, Altmüller J, Koppelman GH, Guerra S, Kronenberg F, Lathrop M, Moffatt MF, O'Connor GT, Strachan DP, Postma DS, London SJ, Schindler C, Kogevinas M, Kauffmann F, Jarvis DL, Demenais F, Probst-Hensch NM. Genome-wide Association Study of Lung Function Decline in Adults with and without Asthma. *Journal of Allergy and Clinical Immunology 2012*, 129(5), 1218-1228.

Hersberger M, **Thun GA**, Imboden M, Brandstätter A, Waechter V, Summerer M, Schmid-Grendelmeier P, Bircher A, Rohrer L, Berger W, Russi EW, Rochat T, Kronenberg F, Probst-Hensch N. Association of STR Polymorphisms in *CMA1* and *IL-4* with Asthma and Atopy: the SAPALDIA Cohort. *Human Immunology 2010*, 71(11), 1154-1160.

Curjuric I, Imboden M, Schindler C, Downs SH, Hersberger M, Liu SL, Matyas G, Russi EW, Schwartz J, **Thun GA**, Postma DS, Rochat T, Probst-Hensch NM; SAPALDIA team. *HMOX1* and *GST* Variants Modify Attenuation of FEF25-75% Decline due to PM10 Reduction. *European Respiratory Journal 2010*, 35(3), 505-514.

SYMPOSIA AND CONFERENCE CONTRIBUTIONS

Thun GA, Bischoff-Ferrari H. Care for the Elderly. Swiss Public Health Conference. *Zürich, Switzerland. August 2013*. Session Chair.

Thun GA, Imboden M, Kogevinas M, Berger W, Rochat T, Probst-Hensch NM. Cell Cycle Control Gene Variants Modify the Association between Asthma and Obesity. Swiss Public Health Conference. *Basel, Switzerland. August 2011*. Oral Presentation.

Thun GA, Ferrarotti I, Imboden M, Rochat T, Zorzetto M, Ottaviani S, Russi EW, Luisetti M, Probst-Hensch NM. *SERPINA1* MS and MZ Heterozygotes and Lung Function Decline in the SAPALDIA Cohort Study. Annual Assembly of the Swiss Society of Pneumology. *Interlaken, Switzerland. May 2011*. Oral Presentation.

Thun GA, Kumar A, Imboden M, Curjuric I, Rochat T, Russi EW, Luisetti M, Probst-Hensch NM. Genome-wide Association with Alpha1-Antitrypsin Blood Levels in a Subset of the SAPALDIA Cohort Study. Annual International Genetic Epidemiology Society Conference. *Boston, USA. October 2010*. Poster Presentation.

Thun GA, Ferrarotti I, Imboden M, Rochat T, Zorzetto M, Ottaviani S, Russi EW, Luisetti M, Probst-Hensch NM. *SERPINA1* MS and MZ Heterozygotes and Lung Function Decline in the SAPALDIA Cohort Study. European Respiratory Society Annual Congress. *Barcelona, Spain. September 2010*. Poster Presentation.

Thun GA, Ferrarotti I, Imboden M, Rochat T, Russi EW, Luisetti M, Probst-Hensch NM. *SERPINA1* S- and Z-genotypes, Lung Function Decline and a GWAS on Alpha1-Antitrypsin Blood Levels in SAPALDIA. Mini-Symposium: The Epidemiology of COPD. *University of Pavia, Pavia, Italy. June 2010*. Oral Presentation.

Thun GA, Ferrarotti I, Imboden M, Rochat T, Russi EW, Luisetti M, Probst-Hensch NM. *SERPINA1* S- and Z-genotypes and Lung Function Decline in the SAPALDIA Cohort Study. Mini-Symposium: Research and Development on Proteases and Antiproteases: with a specific emphasis to Alpha1-Antitrypsin. *Hannover Medical School, Hannover, Germany, January 2010*. Oral Presentation.

COURSE LIST

Swiss School of Public Health +

Observational Epidemiological Studies: Advanced Methods for Design and Analysis	09/2009, Basel	2 ECTS
Systematic Reviews and Meta-Analysis: a Practical Approach	03/2010, Bern	0.75 ECTS
Applied Bayesian Statistics in Medical Research and Health-Care Evaluation	03/2010, Bern	1 ECTS

Multilevel Modelling: Analysis of Clustered Data	11/2010, Basel	1 ECTS
Writing a Journal Article and Getting it Published	03/2012, Bern	1 ECTS

University of Basel

Speaking and Writing English: Advanced	Spring 2010	2 ECTS
Biostatistics I	Autumn 2010	4 ECTS
Applied Bioinformatics	Autumn 2011	2 ECTS
Data Analysis in Epidemiology	Autumn 2011	2 ECTS

London School of Hygiene and Tropical Medicine

Postgraduate Diploma in Epidemiology	10/2010 – 06/2011	20 ECTS