EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Granularity of *SERPINA1* alleles by DNA sequencing in CanCOLD

Nisha Gupta, Nathalie Gaudreault, Sébastien Thériault, Pei Zhi Li, Cyndi Henry, Miranda Kirby, François Maltais, Wan Tan, Jean Bourbeau, Yohan Bossé, for the Canadian Respiratory Research Network, for the CanCOLD study group

Please cite this article as: Gupta N, Gaudreault N, Thériault S, *et al*. Granularity of *SERPINA1* alleles by DNA sequencing in CanCOLD. *Eur Respir J* 2020; in press (https://doi.org/10.1183/13993003.00958-2020).

Granularity of *SERPINA1* alleles by DNA sequencing in CanCOLD

Nisha Gupta[1,2], MD; Nathalie Gaudreault[3], BSc; Sébastien Thériault[3], MD; Pei Zhi Li[1], MSc; Cyndi Henry[3], MSc; Miranda Kirby[4], PhD; François Maltais[3], MD; Wan Tan[5], MD; Jean Bourbeau[1,2], MD; Yohan Bossé[3,6], PhD; for the Canadian Respiratory Research Network, for the CanCOLD study group

1) Research Institute of the McGill University Health Centre, Montréal, Québec, Canada.

2) McGill University, Montréal, Québec, Canada.

3) Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Québec, Canada.

4) Ryerson University, Toronto, Ontario, Canada.

5) University of British-Columbia, Vancouver, British-Columbia, Canada.

6) Department of Molecular Medicine, Laval University, Québec, Québec, Canada.

Corresponding author
Yohan Bossé, Ph.D.
Professor, Laval University
Department of Molecular Medicine
Canada Research Chair in Genomics of Heart and Lung Diseases
Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval
Pavillon Marguerite-d'Youville, Y2106
2725 chemin Sainte-Foy
Québec (Québec)
Canada, G1V 4G5
Tel: 418-656-8711 ext. 3725
Fax: 418-656-4940
email: yohan.bosse@criucpq.ulaval.ca

summary

We found that 15.5% of subjects in this Canadian cohort were carriers of at least one deficient allele affecting alpha-1 antitrypsin serum levels, but only genotypes resulting in severe deficiency and MZ heterozygotes were associated with COPD phenotypes.

Running head

DNA sequencing of *SERPINA1* in CanCOLD

Supplementary material

This article has supplementary material.

ABSTRACT

DNA sequencing of the *SERPINA1* gene to detect alpha-1 antitrypsin deficiency (AATD) may provide a better appreciation of the individual and cumulative impact of genetic variants on alpha-1 antitrypsin (AAT) serum levels and COPD phenotypes.

AAT serum level and DNA sequencing of the coding regions of *SERPINA1* were performed in 1,359 participants of the Canadian Cohort Obstructive Lung Disease (CanCOLD) study. Clinical assessment for COPD included questionnaires, pulmonary function testing and computed tomography (CT) imaging. Phenotypes were tested for association with *SERPINA1* genotypes collated into four groups: normal (MM), mild (MS and MI), intermediate (heterozygote MZ, non-S/non-Z/non-I, compound IS, and homozygote SS), and severe (ZZ and SZ) deficiency. Smoking strata and MZ-only analyses were also performed.

Thirty-four genetic variants were identified including 25 missense mutations. Overall, 8.1% of alleles in this Canadian cohort were deficient and 15.5% of 1,359 individuals were carriers of at least one deficient allele. Four AATD subjects were identified and had statistically lower diffusion capacity and greater CT-based emphysema. No COPD phenotypes were associated with mild and intermediate AATD in the overall cohort or stratified by smoking status. MZ heterozygotes had similar CT-based emphysema, but lowered diffusion capacity compared to normal and mild deficiency.

In this Canadian population-based cohort, comprehensive genetic testing for AATD reveals a variety of deficient alleles affecting 15.5% of subjects. COPD phenotype was demonstrated in severe deficiency and MZ heterozygotes. This study shows the feasibility of implementing a diagnostic test for AATD using DNA sequencing in a large cohort.

Abstract word count: 247

Keywords: Alpha-1 antitrypsin deficiency, emphysema, genetics, chronic obstructive pulmonary disease, rare variant

INTRODUCTION

Alpha-1 antitrypsin deficiency (AATD) is an inherited disorder associated with accelerated rate of lung function decline and early onset emphysema [1-4]. AATD is caused by genetic mutations in the *SERPINA1* gene located on chromosome 14, which encodes an antiprotease called alpha-1 antitrypsin (AAT). The prevalence of AATD is estimated at one in 2,000 to 5,000 in North American populations [5, 6]. Previous studies suggest that severe AATD accounts for 1% to 5% of COPD cases [1, 7, 8], but the exact proportion for the COPD population remains to be described.

Diagnostic methods for AATD vary by countries and regions, but usually follow a multi-step testing algorithm that includes quantification of AAT in serum or plasma, protease inhibitor (Pi) phenotyping by isoelectric focusing, and targeted genotyping for the most common mutations (e.g., S and Z) [9-11]. The gold standard to detect the deficiency is direct DNA sequencing, which was historically considered laborious, expensive and not available in all centers [2]. In the current genomic era, these arguments are no longer valid. More research and clinical laboratories are transitioning to DNA sequencing as the method of choice [11-15]. In so doing, there is an increasing number of rare variants being identified, however little is known about their frequencies and clinical impact.

The goals of this study were two-fold: first, to assess the frequencies of AATD alleles in a Canadian population-based cohort of individuals with COPD, at-risk of COPD, and never-smokers free of airway obstruction, and, second, to evaluate the individual and cumulative impact of AATD alleles on COPD phenotypes including lung function and computed tomography (CT)-based emphysema.

METHODS

CanCOLD study

The Canadian Cohort of Obstructive Lung Disease (CanCOLD) is a prospective cohort study built on the Canadian COPD prevalence study "COLD", which evaluated >5,000 subjects (male and female subjects ≥40 years) recruited through a random sampling frame from nine urban and suburban areas in Canada (ClinicalTrials.gov Identifier: NCT00920348) [16]. Sampling for CanCOLD consists of all COPD subjects as well as age- and sex- matched non-COPD peers (postbronchodilator forced expiratory volume in 1 second [FEV_1]/forced vital capacity [FVC] >0.70) from the COLD study. CanCOLD thus comprises two balanced COPD subpopulations (mild and moderate-severe) and two matched non-COPD subpopulations including ever-smokers (for those at risk) and never-smokers (for the control subjects). Assessments included sociodemographic and clinical status questionnaires (including the COPD Assessment Test, the St. George's Respiratory Questionnaire for COPD [SGRQ-C], mMRC dyspnea scale, and Short Form-36), full pulmonary function tests, CT imaging of the thorax, and incremental maximal cardiopulmonary exercise testing. Details about pulmonary function testing and CT imaging are in the supplementary material. Detailed descriptions of the sampling strategy and assessments can be found in the published protocol [16]. A written informed consent was obtained from all subjects and the study was approved by the Institutional Research Ethical Board of each site. For this analysis, only data from the initial visit were used and a final set of 1,359 CanCOLD subjects with DNA available were selected for *SERPINA1* sequencing.

DNA sequencing of *SERPINA1*

The DNA sequences of the coding regions (i.e. exons 2 to 5) of the *SERPINA1* gene were obtained by Sanger sequencing in the same laboratory for all subjects. For identified variants with unknown allelic background, allele-specific PCR (AS-PCR) was performed in order to amplify and sequence each allele independently. Details are in the supplementary material.

In silico functional analysis of genetic variants

The predicting damaging effects of coding non-synonymous variants were evaluated with PolyPhen-2 [17] and SIFT [18]. The deleteriousness of identified genetic variants was also evaluated with the CADD framework [19]. Finally, clinical interpretation was queried in ClinVar [20].

Measurement of AAT serum levels

Blood samples from CanCOLD participants were drawn by venipuncture and processed within 2 hours of collection. Serum aliquots were stored at -80°C until analysis. AAT serum levels were measured by immunoturbidimetry on a COBAS INTEGRA 800 analyser (Roche Diagnostics, Laval, Canada). The coefficient of variation for intermediate precision was estimated at 2.8% at a level of 0.76 g/L and 2.2% at a level of 1.79 g/L.

Statistical analysis

The effects of genetic variants on AAT serum levels were evaluated using Wald tests for quantitative phenotype as implemented in PLINK [21]. Clinical data were compared across groups using ANOVA for continuous variables and chi-square for categorical variables. Analyses were also carried out stratified by smoking status (never- and ever-smokers). Analyses were performed with and without adjustment for covariates including age, sex and study sites. Post-hoc Tukey multiple pairwise-comparison tests were used for statistically significant ANOVA to identify groups that differed.

RESULTS

Identified genetic variants in CanCOLD

Table S1 shows the clinical characteristics of the 1,359 CanCOLD participants evaluated in this study. DNA sequencing of exons 2 to 5 in these individuals identified 34 genetic variants. The location of these variants relative to the intron-exon structure of the *SERPINA1* gene is illustrated in **Figure 1**. **Table 1** shows the variants based on standardized genetic, nucleotide, protein, and Pi typing nomenclatures. For each variant, the pathogenicity scores and the allele frequency among the 1,359 CanCOLD subjects are also provided in **Table 1**. Three new variants were found including two in exon 3 (Lys222Lys and Lys274Asn) and one in exon 4 (Leu318Phe) (**Figure 1**). These new variants are rare with allele frequencies of 0.04% (one heterozygote subject) and predict benign or inconsistent across pathogenicity scores (**Table 1**).

Allele and genotype calling

DNA sequencing provides more granularity compared to the conventional Pi system. We thus had to use a refined strategy to call alleles and genotypes. Humans are diploid organism, meaning that there are two alleles per individual. For each allele, one or more genetic variants can be present. The absence or presence of a genetic variant is called in relation to an allele (DNA sequence) of reference. For *SERPINA1*, the allele of reference and the most common in human is labeled M1 (Val^{213}) using the Pi system. In CanCOLD, for example, 1,246 out of 2,718 alleles (1,359 individuals x 2 alleles = 2,718 alleles) are on this background allele. Accordingly, we called the 2,718 alleles and the 1,359 genotypes of CanCOLD participants based on the absence or presence of the 34 genetic variants identified (**Figure 1** and **Table 1**). **Figure 2** illustrates five representative individuals to better conceptualize the allele and genotype calling process.

Frequency of alleles

The 34 genetic variants in CanCOLD results in 40 distinct alleles. **Table 2** shows the distribution of alleles in the overall CanCOLD cohort and by recruitment sites. The most frequent normal alleles in

this population are 45.8% for M1 (Val^{213}), 17.5% for M1 (Ala^{213}), 14.6% for M2, 9.1% for M3, 1.8% for $M3_{Riedenburg}$, and 1.3% for M4. In total, 91.9% of the 2,718 alleles are normal. In contrast, the others are pathogenic including known deficient or dysfunctional alleles such as S (5.7%), Z (1.7%), F (0.2%), I (0.2%) as well as more rare variants such as P_{Lowell} found in two heterozygote subjects and $M_{Procida}$ and $M_{Wurzburg}$ each found in one heterozygote subject. Four additional rare variants, each found in a single individual, are claimed pathogenic by all scores (**Table 1**) including M3.Pro255Thr, $S_{Donosti}$, S_{Munich}, and S.Arg101Cys. Taking together, 220 deficient alleles are identified and indicate that 8.1% of alleles (220 out of 2,718) in this Canadian cohort are deficient. The percentage of deficient alleles varies from 5.5% to 11% across recruitment sites (**Table 2**).

Genotype frequencies

The genotype of each individual consists of the combination of both alleles. **Figure 3** illustrates the number of individuals in any combinations of normal and deficient alleles. Two subjects carry two copies of the Z allele and are thus confirmed cases of AATD. One other subject was compound heterozygote SZ. A second SZ individual was identified, but carrying also the Ser14Phe mutation. Allele-specific PCR indicated that this mutation occurs on the background of S allele, which is consistent with a previously reported allele labelled $S_{Donosti}$ [14]. As a result, there were a total of four CanCOLD subjects with severe AATD. Forty additional individuals were heterozygotes for the Z allele. Five SS individuals were found, and 143 additional individuals carried the S allele (142 MS and 1 IS). Six subjects were carriers of the F allele. Six carriers of allele I were also found including one IS. Rarer deficient alleles were identified in 2 P_{Lowell}, 1 $M_{Procida}$, 1 $M_{Wurzburg}$, 1 S_{Munich}, and 1 M3.Pro255Thr heterozygote carriers. Overall, 210 out of 1,359 subjects (15.5%) were carriers of at least one deficiency allele.

***SERPINA1* genotyping grouping**

SERPINA1 genotyping groups were then generated from the DNA sequencing data to reflect the resulting AAT serum level or functional activity of AAT (**Figure 4**). At the most detailed level, the 40 distinct alleles found in CanCOLD lead to 77 distinct genotypes with frequencies ranging from 0.07% (1 out of 1,359 individuals) to 21.3% (289 out 1,359 individuals, i.e. genotype M1M1) (see x-axis of **Figure 4**). Although there is no standard for grouping genotypes, we have classified individuals into four groups based on the information of both alleles and the known or predicting effect of each genetic variant on AAT levels or activity. Group 1 (no deficiency) includes all individuals with two normal alleles (n=1,149). Group 2 (mild deficiency) are individuals with one normal allele and one S or I allele (n=147, MS and MI). Group 3 (intermediate deficiency) consists of individuals that are heterozygotes (MZ, MF, $MM_{Procida}$, $MM_{Wurzburg}$, MP_{Lowell}, MS_{Munich}, MS.Arg101Cys, MM3.Pro255Thr), homozygotes for the S allele (SS) and compound heterozygotes for deficient alleles (IS) (n=59). Group 4 (severe deficiency) are AATD individuals (n=4, ZZ, SZ, and $S_{Donosti}Z$).

Association of *SERPINA1* genotyping groups with AAT serum levels and COPD phenotypes

Table 3 shows the clinical characteristics of CanCOLD participants according to the *SERPINA1* genotyping groups. Serum AAT levels were strongly associated with genotyping groups (ANOVA p value <0.001, **Figure 5A**). The 6-heterozygote individuals for the F allele have serum AAT levels ranging with those with two normal alleles. Diffusion capacity to carbon monoxide (DLCO) and the percentage of lung voxels below -950 Hounsfield units (LAA-950) were also statistically different across these groups (**Table 3** & **Figure 5B-C**). These associations were driven by AATD individuals (group 4: severe deficiency) as no significant differences were observed between group 1 (no deficiency) and group 3 (intermediate deficiency). No statistically significant differences were observed across these genotyping groups for smoking habits, lung function, symptoms and quality of life, self-reported comorbidities, and respiratory medications (**Table 3**).

Intermediate AAT deficiency, mostly defined by MZ heterozygotes, was previously associated with an increased risk of COPD, especially in the presence of smoking exposure [22-25]. We thus repeated the analyses by smoking status (never- and ever-smokers). As observed for the entire cohort, the *SERPINA1* genotyping groups were significantly different for LAA-950 in both never- and ever-smokers, and DLCO in never-smokers (**Table S2**). Again, post-hoc tests revealed that group 4 (severe deficiency), but not group 3 (intermediate deficiency), is significantly different from group 1 (no deficiency). We have also evaluated the interaction between *SERPINA1* genotyping groups and smoking (ever vs. never smokers) on COPD phenotypes, but no significant interactions were identified (**Table 3**).

Analysis of MZ heterozygotes

Finally, COPD phenotypes were compared by *SERPINA1* genotyping groups, but this time limiting intermediate deficiency to MZ heterozygotes (n=40). Compared to normal and mild deficiency, MZ had similar lung function and CT-based emphysema (**Table S3**). However, DLCO was statistically lower ($P<0.001$) in MZ (95.5% ± 21.2%) and severe deficiency (69.4% ± 19.1%) compared to normal (106.4% ± 24.3%) and mild deficiency (109.9% ± 26.5%).

DISCUSSION

In this Canadian cohort, 34 genetic variants were identified by sequencing the coding regions of the *SERPINA1* gene. This includes genetic variants accounting for common normal alleles such as M1, M1 (Ala^{213}), M2, M3 and M4, but also more rare normal alleles such as $M3_{Riedenburg}$, $M2_{Obernburg}$, V, $M6_{Passau}$, and $M1_{Cadiz}$. Collectively, 91.9% of all the alleles evaluated were considered normal. In contrast, a variety of 220 deficient alleles was found in these 1,359 individuals. Not surprisingly, the most common were S and Z, but mutations causing alleles I, F, P_{Lowell}, $M_{Procida}$, $M_{Wurzburg}$,

M3.Pro255Thr, $S_{Donosti}$, S_{Munich}, and S.Arg101Cys were also identified. Deficient alleles were observed in 210 individuals, indicating that 15.5% of CanCOLD subjects carried at least one deficient allele. Grouping subjects by genotypes reflecting the serum level or functional activity of AAT indicated that severe deficiency (ZZ, SZ, and $S_{Donosti}Z$) had lower diffusion capacity and greater CT-based emphysema compared with subjects with no deficiency (MM). These differences were not observed in mild (MS and MI) and intermediate (heterozygote MZ, heterozygote non-S/non-Z/non-I, homozygote SS, and compound heterozygote IS) deficiency. However, in MZ-only analysis, MZ heterozygotes had lower diffusion capacity compared to normal and mild deficiency.

Four subjects (0.29%) had AATD in CanCOLD, i.e. ZZ, SZ, or $S_{Donosti}Z$. This is lower than the 1.9% of ZZ found in 965 severe COPD patients [7], but higher than the frequency of severe AATD commonly reported at 1 out of 2,000-5,000 individuals [5, 6]. Random sampling in this study [16] which includes subgroups without airway obstruction, those at-risk of COPD, and individuals with COPD with a range of severity may explain this difference. Nevertheless, deficient alleles together including Z, S, I, F, P_{Lowell}, $M_{Procida}$, $M_{Wurzburg}$, M3.Pro255Thr, $S_{Donosti}$, S_{Munich}, and S.Arg101Cys account for 8.1% of all alleles tested (220 out of 2,718). Overall, these alleles affect 15.5% of individuals, which may have a genetic predisposition to develop COPD [22, 23]. Our results highlight that less frequent deficient alleles, beyond S and Z, affects 20 individuals (1.5%) in CanCOLD.

In this study, we identified 59 individuals with intermediate AAT deficiency. The novelty of our study is that we have included genotypes associated with intermediate deficiency beyond MZ including heterozygote non-Z (MF, $MM_{Procida}$, $MM_{Wurzburg}$, MP_{Lowell}, MS_{Munich}, MS.Arg101Cys, MM3.Pro255Thr), homozygote SS, and compound heterozygote IS. However, the COPD phenotypes of this group was not statistically different compared to individuals with no deficiency (MM), which was consistent across smoking status strata. Sample size may have limited our ability to detect the

effect of intermediate deficiency on COPD phenotypes. Alternatively, the ascertainment method based on random sampling in CanCOLD may also explain this lack of association. For many COPD cohorts, ascertainment is based on patients referred specifically for obstructive airway disease. In the natural history of AATD, we know that index cases, those seen by medical care facilities, are more ill compared to nonindex cases (family members) [26]. In a population-based cohort like CanCOLD, fewer individuals have developed pulmonary symptoms and associated inflammatory milieu that would favor disease progression in the intermediate AAT deficiency state. Finally, lower smoking exposure may also have limited expressivity.

SERPINA1 genotyping grouping used in this study cannot be directly compared with previous studies. To make a fair comparison, we have performed MZ-only analysis and showed lower diffusing capacity in MZ heterozygotes. This result seems consistent with previous studies that have showed that MZ individuals are characterized by lower FEV_1/FVC ratio and more radiographic emphysema [24], accelerated FEF25-75% decline in the presence of smoking or obesity [25], lower lung function and CT-based emphysema [22], and lower lung function and greater COPD risk, especially in ever-smokers [23].

This study demonstrates that DNA sequencing is feasible in a large cohort. DNA sequencing provides a complete assessment of the gene causing AATD, namely *SERPINA1*. It can detect conventional Pi deficient alleles (e.g., Z and S), but also rare and novel genetic variants. Diagnosis at the DNA level is the modern gold standard information needed for hereditary disorders. Importantly, DNA sequencing can be implemented in any laboratory with a standard DNA sequencing system, has a fast turnaround time, and may eventually replace the historical multistep testing algorithms to detect AATD used in most countries [9-11]. With established consistency in diagnosing AATD at the DNA level, there is

potential to expedite the clinical decision making process recommended by many national and international lung societies for targeted testing and augmentation therapy [1-4].

In this study, we have not performed cost-effectiveness analysis of *SERPINA1* DNA sequencing relative to current multistep testing algorithms to detect AATD. For practical considerations, we provide costs for DNA sequencing. In Canadian currency, the cost of materials and supplies to sequence the 1,359 CanCOLD participants is estimated at 31,500$ [(1.85$ per PCR reaction + 3.75$ per sequencing reaction) x 4 amplicons per sample x 1,359 samples]. About 10$ per sample must be added for DNA extraction. An effective workforce and the number of samples processed at the same time are key to keep the cost low. It takes about 1½ day to complete the protocol. The salary of a laboratory technician for that time is thus the main factor fixing the cost per sample. In this research project, we were able to process samples in 96-well plate format and save costs. Lower throughput is expected in a realistic clinical health care system. For example, in our laboratory, the overall cost of *SERPINA1* DNA sequencing per sample is 650$, which drops to 150$ when 8 samples are processed at the same time. This implies that all equipment are in place, including a sequencer and about 20,000$ of ancillary equipment (thermocycler, centrifuges, pipettes, water bath, vortex, and electrophoresis system). DNA sequencers come in different throughput formats and prices, but one that can process 4 samples at the time is available at approximately 50,000$. Sequencing reactions can also be outsourced if a sequencer is not available. All these costs must be balanced with the resulting benefits. Here, we tested 1,359 individuals and obtained a conclusive AATD diagnostic for all of them, i.e. success rate = 100%. The single step test can be performed in 2 days (more realistically offered within a week) and thus has the potential to end the diagnostic odyssey of patients with AATD.

This study has many strengths, including being the first Canadian COPD cohort that has recruited its participants from the general population rather than convenient sampling often utilized in clinical

research. Furthermore, the research participants have been extensively phenotyped for COPD (clinical, physiology, CT scan with qualitative and quantitative evaluation). The limitations include the relatively limited sample size. As with other cohort studies, participation bias may affect the representativeness of the cohort relative to the underlying population. The current study has been limited to cross-sectional data. Longitudinal data collection is ongoing, and future analyses are needed with respect to COPD development and progression. We attempted to classify all 34 genetic variants identified in CanCOLD participants as normal or deficient (**Table 1**). This was based on all possible sources of information including previous literature, location of the variants in the protein structure, pathogenicity scores, and AAT serum levels among carriers. However, some uncertainties remain about the classification of rare variants and further functional studies will be needed. Finally, the 4-group classification scheme used in this study (no, mild, intermediate and severe deficiency) that reflects the combination of the two alleles observed in each individual is likely to be challenged and refined with time. However, grouping genotypes has always been inherent (consciously or not) in the field of AATD. For example, how many individuals would be MM in CanCOLD based on the conventional Pi system? With DNA sequencing, we demonstrated that the no deficiency group (individuals with two normal alleles) is in fact a collection of 49 different genotypes (see x-axis of **Figure 4**). Groups were thus formed in this study using an unprecedented level of granularity. The strong association and progressive decline in AAT levels with *SERPINA1* genotyping groups (**Figure 5A**) support our grouping scheme.

In conclusion, this study identified 34 genetic variants in the *SERPINA1* gene resulting in 220 deficient alleles from 1,359 individuals. Four AATD cases were identified and the remaining deficient alleles were carried by 206 individuals, indicating that deficient alleles affected 15.5% of subjects in this Canadian population-based cohort. Patients with genotypes resulting in severe AAT deficiency (AATD cases) are more susceptible to develop airway obstruction as demonstrated by lower diffusion capacity and greater CT-based emphysema. In contrast, mild and intermediate deficiency genotypes were not

associated with COPD phenotypes in this random sampling cohort except a lower diffusion capacity in MZ heterozygotes. Overall, this study demonstrates the frequency of deficient alleles beyond the most common S and Z alleles in a population-based cohort, and the feasibility of DNA sequencing on a large-scale to provide an accurate and definitive diagnosis for AATD. The longitudinal evaluation will be important to determine the effects of *SERPINA1* genotypes on COPD risk and progression.

Acknowledgments

The authors thank the men and women who participated in the study and individuals in the CanCOLD Collaborative research Group:

Executive Committee: Jean Bourbeau (McGill University, Montreal, QC, Canada); Wan C Tan, J Mark FitzGerald, Don D Sin (UBC, Vancouver, BC, Canada); Darcy D Marciniuk (University of Saskatoon, Saskatoon, SK, Canada); Denis E O'Donnell (Queen's University, Kingston, ON, Canada); Paul Hernandez (University of Halifax, Halifax, NS, Canada); Kenneth R Chapman (University of Toronto, Toronto, ON, Canada); Robert Cowie (University of Calgary, Calgary, AB, Canada); Shawn Aaron (University of Ottawa, Ottawa, ON, Canada); François Maltais (University of Laval, Quebec City, QC, Canada).

International Advisory Board: Jonathon Samet (the Keck School of Medicine of USC, California, USA); Milo Puhan (John Hopkins School of Public Health, Baltimore, USA); Qutayba Hamid (McGill University, Montreal, QC, Canada); James C Hogg (University of British Columbia, James Hogg Research Center, Vancouver, BC, Canada).

Operations Center: Jean Bourbeau (PI), Carole Baglole, Carole Jabet, Palmina Mancino, Yvan Fortier (University of McGill, Montreal, QC, Canada); Wan C Tan (co-PI), Don Sin, Sheena Tam, Jeremy Road, Joe Comeau, Adrian Png, Harvey Coxson, Miranda Kirby, Jonathon Leipsic, Cameron Hague (University of British Columbia, James Hogg Research Center, Vancouver, BC, Canada).

Economic Core: Mohsen Sadatsafavi (University of British Columbia, Vancouver, BC).

Public Health Core: Teresa To, Andrea Gershon (University of Toronto).

Data management and Quality Control: Wan C Tan, Harvey Coxson (UBC, Vancouver, BC, Canada); Jean Bourbeau, Pei-Zhi Li, Jean-Francois Duquette, Yvan Fortier, Andrea Benedetti, Denis Jensen (McGill University, Montreal, QC, Canada); Denis O'Donnell (Queen's University, Kingston, ON, Canada.

Field Centers: Wan C Tan (PI), Christine Lo, Sarah Cheng, Elena Un, Cindy Fung, Wen Tiang Wang, Faize Faroon, Olga Radivojevic, Carl Zou (UBC James Hogg Research Center, Vancouver, BC, Canada); Jean Bourbeau (PI), Palmina Mancino, David Latreille, Jacinthe Baril, Laura Labonte (McGill University, Montreal, QC, Canada); Kenneth Chapman (PI), Patricia McClean, Nadeen Audisho (University of Toronto, Toronto, ON, Canada); Brandie Walker, Robert Cowie (PI), Ann Cowie, Curtis Dumonceaux, Lisette Machado (University of Calgary, Calgary, AB, Canada); Paul Hernandez (PI), Scott Fulton, Kristen Osterling (University of Halifax, Halifax, NS, Canada); Shawn Aaron (PI), Kathy Vandemheen, Gay Pratt, Amanda Bergeron (University of Ottawa, Ottawa, ON, Canada); Denis O'Donnell (PI), Matthew McNeil, Kate Whelan (Queen's University, Kingston, ON, Canada); François Maltais (PI), Cynthia Brouillard (University of Laval, Quebec City, QC, Canada); Darcy Marciniuk (PI), Ron Clemens, Janet Baran (University of Saskatoon, Saskatoon, SK, Canada).

Author contributions

Y.B. and J.B. conceived the study, have full access to the data, and take responsibility for the integrity of the data and article. N.Gaudreault and C.H. performed the DNA sequencing. S.T. measured the AAT serum levels. M.K., F.M., W.T. and J.B. contributed to the study design, performed clinical recruitment and phenotyping of patients. P.Z.L. and Y.B. performed data analysis. N.Gupta, J.B. and Y.B. wrote the manuscript. N.Gaudreault, S.T., M.K., F.M. and W.T. edited the manuscript for intellectual content. All authors approved the final manuscript.

Financial support

This work was supported by Grifols Canada Ltd., the *Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec*, the Respiratory Health Network of the *Fonds de recherche Québec – Santé* (FRQS), and the Canadian Institutes of Health Research (MOP - 123369). The Canadian Cohort Obstructive Lung Disease (CanCOLD) study is currently funded by the Canadian Respiratory Research Network (CRRN); industry partners: Astra Zeneca Canada Ltd; Boehringer Ingelheim Canada Ltd; GlaxoSmithKline Canada Ltd; Novartis. Researchers at RI-MUHC Montreal and Icapture Centre Vancouver lead the project. Previous funding partners are the CIHR (CIHR/ Rx&D Collaborative Research Program Operating Grants- 93326); the RHN of the FRQS; industry partners: Almirall; Merck Nycomed; Pfizer Canada Ltd; and Theratechnologies. S Thériault holds a Junior 1 Clinical Research Scholar award from the FRQS. M Kirby holds a Canada Research Chair in Quantitative Imaging. F Maltais holds a GSK Research Chair on COPD at Université Laval. J Bourbeau holds a GSK/CIHR Research Chair on COPD at McGill University. Y Bossé holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

Conflict of interest

MK is a consultant for VIDA Diagnostics Inc.

References

1. 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* 2003: 168(7): 818-900.
2. Marciniuk DD, Hernandez P, Balter M, Bourbeau J, Chapman KR, Ford GT, Lauzon JL, Maltais F, O'Donnell DE, Goodridge D, Strange C, Cave AJ, Curren K, Muthuri S, Canadian Thoracic Society CCAA-ADEWG. Alpha-1 antitrypsin deficiency targeted testing and augmentation therapy: a Canadian Thoracic Society clinical practice guideline. *Can Respir J* 2012: 19(2): 109-116.
3. Miravitlles M, Dirksen A, Ferrarotti I, Koblizek V, Lange P, Mahadeva R, McElvaney NG, Parr D, Piitulainen E, Roche N, Stolk J, Thabut G, Turner A, Vogelmeier C, Stockley RA. European Respiratory Society statement: diagnosis and treatment of pulmonary disease in alpha1-antitrypsin deficiency. *Eur Respir J* 2017: 50(5).
4. Alpha 1-antitrypsin deficiency: memorandum from a WHO meeting. *Bull World Health Organ* 1997: 75(5): 397-415.
5. de Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. *Chest* 2002: 122(5): 1818-1829.
6. Luisetti M, Seersholm N. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. *Thorax* 2004: 59(2): 164-169.
7. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. *Chest* 1986: 89(3): 370-373.
8. Abboud RT, Ford GT, Chapman KR. Alpha1-antitrypsin deficiency: a position statement of the Canadian Thoracic Society. *Can Respir J* 2001: 8(2): 81-88.
9. Snyder MR, Katzmann JA, Butz ML, Wiley C, Yang P, Dawson DB, Halling KC, Highsmith WE, Thibodeau SN. Diagnosis of alpha-1-antitrypsin deficiency: An algorithm of quantification, genotyping, and phenotyping. *Clin Chem* 2006: 52(12): 2236-2242.
10. Ferrarotti I, Scabini R, Campo I, Ottaviani S, Zorzetto M, Gorrini M, Luisetti M. Laboratory diagnosis of alpha1-antitrypsin deficiency. *Transl Res* 2007: 150(5): 267-274.
11. Graham RP, Dina MA, Howe SC, Butz ML, Willkomm KS, Murray DL, Snyder MR, Rumilla KM, Halling KC, Highsmith WE, Jr. SERPINA1 Full-Gene Sequencing Identifies Rare Mutations Not Detected in Targeted Mutation Analysis. *J Mol Diagn* 2015: 17(6): 689-694.
12. Maltais F, Gaudreault N, Racine C, Theriault S, Bossé Y. Clinical Experience with SERPINA1 DNA Sequencing to Detect Alpha-1 Antitrypsin Deficiency. *Ann Am Thorac Soc* 2018: 15(2): 266-268.
13. Silva D, Oliveira MJ, Guimaraes M, Lima R, Gomes S, Seixas S. Alpha-1-antitrypsin (SERPINA1) mutation spectrum: Three novel variants and haplotype characterization of rare deficiency alleles identified in Portugal. *Respir Med* 2016: 116: 8-18.
14. Matamala N, Lara B, Gomez-Mariano G, Martinez S, Retana D, Fernandez T, Silvestre RA, Belmonte I, Rodriguez-Frias F, Vilar M, Saez R, Iturbe I, Castillo S, Molina-Molina M, Texido A, Tirado-Conde G, Lopez-Campos JL, Posada M, Blanco I, Janciauskiene S, Martinez-Delgado B. Characterization of Novel Missense Variants of SERPINA1 Gene Causing Alpha-1 Antitrypsin Deficiency. *Am J Respir Cell Mol Biol* 2018: 58(6): 706-716.
15. Ortega VE, Li X, O'Neal WK, Lackey L, Ampleford E, Hawkins GA, Grayeski PJ, Laederach A, Barjaktarevic I, Barr RG, Cooper C, Couper D, Han MK, Kanner RE, Kleerup EC, Martinez FJ, Paine Iii R, Peters SP, Pirozzi C, Rennard SI, Woodruff PG, Hoffman EA, Meyers DA, Bleecker ER, Subpopulations N, Intermediate Outcomes Measures in CS. The Effects of Rare SERPINA1 Variants on Lung Function and Emphysema in SPIROMICS. *Am J Respir Crit Care Med* 2019.
16. Bourbeau J, Tan WC, Benedetti A, Aaron SD, Chapman KR, Coxson HO, Cowie R, Fitzgerald M, Goldstein R, Hernandez P, Leipsic J, Maltais F, Marciniuk D, O'Donnell D, Sin DD, Cancold Study

G. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD. *COPD* 2014: 11(2): 125-132.
17. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010: 7(4): 248-249.
18. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009: 4(7): 1073-1081.
19. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014: 46(3): 310-315.
20. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018: 46(D1): D1062-D1067.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007: 81(3): 559-575.
22. Foreman MG, Wilson C, DeMeo DL, Hersh CP, Beaty TH, Cho MH, Ziniti J, Curran-Everett D, Criner G, Hokanson JE, Brantly M, Rouhani FN, Sandhaus RA, Crapo JD, Silverman EK, Genetic Epidemiology of CI. Alpha-1 Antitrypsin PiMZ Genotype Is Associated with Chronic Obstructive Pulmonary Disease in Two Racial Groups. *Ann Am Thorac Soc* 2017: 14(8): 1280-1287.
23. Molloy K, Hersh CP, Morris VB, Carroll TP, O'Connor CA, Lasky-Su JA, Greene CM, O'Neill SJ, Silverman EK, McElvaney NG. Clarification of the risk of chronic obstructive pulmonary disease in alpha1-antitrypsin deficiency PiMZ heterozygotes. *Am J Respir Crit Care Med* 2014: 189(4): 419-427.
24. Sorheim IC, Bakke P, Gulsvik A, Pillai SG, Johannessen A, Gaarder PI, Campbell EJ, Agusti A, Calverley PM, Donner CF, Make BJ, Rennard SI, Vestbo J, Wouters EF, Pare PD, Levy RD, Coxson HO, Lomas DA, Hersh CP, Silverman EK. alpha(1)-Antitrypsin protease inhibitor MZ heterozygosity is associated with airflow obstruction in two large cohorts. *Chest* 2010: 138(5): 1125-1132.
25. 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.
26. Stoller JK. The National Institutes of Health/National Heart, Lung, and Blood Institute Registry for patients with severe deficiency of alpha 1-antitrypsin. *In:* Crystal RG, ed. Alpha 1-antitrypsin deficiency: Biology, pathogenesis, clinical manifestations, therapy. Taylor & Francis Group, New York, 1996.
27. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000: 15(1): 7-12.
28. Ringenbach MR, Banta E, Snyder MR, Craig TJ, Ishmael FT. A challenging diagnosis of alpha-1-antitrypsin deficiency: identification of a patient with a novel F/Null phenotype. *Allergy Asthma Clin Immunol* 2011: 7(1): 18.

Figure legends

Figure 1. The exon-intron structure of the *SERPINA1* gene and the localization of the identified genetic variants. The coding exons are shown in black and the untranslated regions in grey. Genetic variants are illustrated with their rs numbers (if available), protein nomenclature, and genotyping counts in parentheses for 1,359 individuals. Newly identified variants are illustrated in green and named based on standard gene mutation nomenclature [27]. Variants causing a change in conventional Pi typing are illustrated in blue and red for normal and deficient alleles, respectively. Note that rs112030253 in exon 2 is indicated twice as it is a multi-allelic polymorphism and three alleles are observed in the CanCOLD cohort.

Figure 2. Schematic examples of allele and genotype calling derived from DNA sequencing data for five individuals. **Case 1** is homozygote for the M1 (Val^{213}) allele. This is the most common and normal allele in human. The DNA nucleotide sequence of M1 is thus considered the reference. **Case 2** is a typical MZ individual with one normal M1 allele and one deficient Z allele. The Z allele is the result of a missense mutation (Glu342Lys) in exon 5. Except for very rare cases, the Z allele is found on the M1 (Ala^{213}) background. **Case 3** is a SZ individual with one Z allele and one S allele characterized by the Glu264Val mutation in exon 3. **Case 4** is carrier of the rare mutation Arg101Cys on the S background. Arg101Cys is considered deleterious based on pathogenicity scores (Table 1). There are thus two hits (Cys101 and Val264) on the same allele. For this individual, knowing the allelic background was clinically relevant to determine the AATD status, i.e. the individual is heterozygote for one deficient allele and not compound heterozygote (two deficient alleles). **Case 5** is carrier of a typical M3 allele characterized by Glu376Asp and a second unusual allele with His101 and Ala213, labelled R for rare allele background. These are all considered normal genetic variants, so there is no clinical implication

of determining the allelic background of this individual. However, the observation is of great interest for population genetics. Note that this allele and genotype calling process was used to derive all alleles (n=2,718) and genotypes (n=1,359) of the 1,359 CanCOLD individuals.

Figure 3. Number of patients identified in any pairwise combinations of normal and deficient alleles. Alleles are annotated based on the legacy nomenclature, i.e. protease inhibitor (Pi) system, if available, or on protein nomenclature. Normal alleles without Pi naming were combined with their respective background to ease visualization (for example, allele M1.Val302Ile was merged with M1, this allele does not have a Pi naming and Val302Ile is benign). Each individual is represented in only one cell and the combinations of alleles for the 1,359 CanCOLD participants are indicated. The color of the squares illustrates the expected level of deficiency on a black and white scale where white indicates no deficiency and black indicates severe deficiency. *Rare allelic background (1. His101 and Ala213 on the same allele, and 2. His101, Ala213, and Val340 on the same allele).

Figure 4. Serum levels of AAT by *SERPINA1* genotypes. *SERPINA1* genotyping groups are depicted by rectangles. Groups were generated to reflect the resulting impact of the genetic variants on AAT serum level or functional activity of AAT. For rare mutations of unknown biological and clinical significance, the grouping was guided based on the actual serum level of AAT among carriers and pathogenicity scores. This classification system results in four groups: Group 1 (no deficiency) includes all individuals with two normal alleles (n=1,149). Group 2 (mild deficiency) are individuals with one normal allele and one S or I allele (n=147, MS and MI). Group 3 (intermediate deficiency) consists of individuals that are heterozygotes (MZ, MF, $MM_{Procida}$, $MM_{Wurzburg}$, MP_{Lowell}, MS_{Munich}, MS.Arg101Cys, MM3.Pro255Thr), homozygotes for the S allele (SS) and compound heterozygotes for

deficient alleles (IS) (n=59). Group 4 (severe deficiency) are AATD individuals (n=4, ZZ, SZ, and $S_{Donosti}Z$). AAT serum levels are not available for 15 out of 1,359 individuals including the single individual carrying the $M1_{Cadiz}$ allele, which explains the absence of boxplot for the M1(Cadiz)_M3 genotype. M3_R, "R" indicates rare allelic background.

Figure 5. Serum levels of AAT (A), DLCO (B), and LAA-950 (C) by *SERPINA1* genotyping groups derived from DNA sequencing data. Genotyping groups are on the x-axis. Group 1 includes all individuals with two normal alleles (n=1,149). Group 2 are individuals with one normal allele and one S or I allele (n=147, MS and MI). Group 3 consists of individuals that are heterozygotes (MZ, MF, $MM_{Procida}$, $MM_{Wurzburg}$, MP_{Lowell}, MS_{Munich}, MS.Arg101Cys, MM3.Pro255Thr), homozygotes for the S allele (SS) and compound heterozygotes for deficient alleles (IS) (n=59). Group 4 are AATD individuals (n=4, ZZ, SZ, and $S_{Donosti}Z$). Boxplot boundaries represent the first and third quartiles, whiskers are the most extreme data point which is no more than 1.5 times the interquartile range, and the center mark represents the median. Red dots illustrate the six individuals carrying one F allele. F allele have dysfunctional proteins but normal AAT blood levels [28], which exemplifies the intrinsic limitation of relying solely on serum measurement of AAT and commercially available genotyping tests to establish the diagnosis of AATD.

Table 1. Nomenclature of the 34 genetic variants identified, allele frequencies in CanCOLD, and pathogenicity scores

Exon	rs #	Nucleotide	Protein	Pi typing	MAF	Effect	AAT (p value)*	PolyPhen-2[†]	SIFT[‡]	CADD	ClinVar
2	rs1343069141	c.18G>A	Ser-19Ser		0.04%	Unknown	0.789	NA	NA	0.004	NA
2	rs147283849	c.43C>T	Leu-10Leu		0.11%	Unknown	0.068	NA	NA	10.97	Conflicting
2	rs1379209512	c.68T>A	Leu-2Gln		0.07%	Unknown	0.994	0.92 (P)	0.001 (D)	10.56	NA
2	rs745463238	c.113C>T	Ser14Phe	$S_{Donosti}$	0.04%	Deficient	**0.001**	0.708 (P)	0.011 (D)	15.18	NA
2	rs150784949	c.171C>T	Phe33Phe		0.07%	Unknown	0.482	NA	NA	0.991	Conflicting
2	rs28931570	c.187C>T	Arg39Cys	I	0.22%	Deficient	**0.026**	1.0 (D)	0.0 (D)	24.4	Pathogenic
2	rs28931569	c.194T>C	Leu41Pro	$M_{Procida}$	0.04%	Deficient	0.100	0.995 (D)	0.002 (D)	17.59	Pathogenic/Likely pathogenic
2	rs113817720	c.244G>A	Ala58Thr		0.04%	Unknown	0.441	0.276 (B)	0.04 (D)	18.76	Uncertain significance
2	rs111850950	c.250G>A	Ala60Thr	$M6_{Passau}$	0.07%	Normal	0.355	0.999 (D)	0.057 (T)	26.3	Uncertain significance
2	rs758820515	c.335C>T	Pro88Leu		0.04%	Unknown	0.172	0.628 (P)	0.123 (T)	8.747	NA
2	rs766025736	c.373C>T	Arg101Cys		0.04%	Unknown	0.154	0.969 (D)	0.003 (D)	22.3	NA
2	rs709932	c.374G>A	Arg101His	M4 / M2	16.00%	Normal	**0.006**	0.0 (B)	1.0 (T)	0.322	Benign/Likely benign
2	rs1344951022	c.411C>G	Thr113Thr		0.04%	Unknown	0.892	NA	NA	6.218	NA
2	rs20546	c.424C>T	Leu118Leu	$M3_{Riedenburg}$	1.77%	Normal	0.358	NA	NA	3.036	Benign/Likely benign
2	rs112030253	c.514G>T	Gly148Trp	$M2_{Obernburg}$	0.30%	Normal	0.838	0.973 (D)	0.027 (D)	13.59	Conflicting
2	rs112030253	c.514G>A	Gly148Arg	V	0.11%	Normal	0.675	0.0 (B)	0.934 (T)	0.008	Conflicting
2	rs149770048	c.523G>A	Glu151Lys	$M1_{Cadiz}$	0.04%	Normal	NA	0.05 (B)	0.097 (T)	9.991	NA
2	rs200414579	c.573G>A	Gly167Gly		0.04%	Unknown	0.564	NA	NA	0.277	NA
3	rs6647	c.710T>C	Val213Ala	M1	20.27%	Normal	**0.0009**	0.0 (B)	0.597 (T)	0.002	Benign/Likely benign
3	New1	c.738G>A	Lys222Lys		0.04%	Unknown	0.803	NA	NA	0.154	NA
3	rs28929470	c.739C>T	Arg223Cys	F	0.22%	Deficient	0.650	0.522 (P)	0.001 (D)	23.1	Pathogenic
3	rs759736224	c.835C>A	Pro255Thr		0.04%	Unknown	0.586	1.0 (D)	0.0 (D)	23.8	NA
3	rs121912714	c.839A>T	Asp256Val	P_{Lowell}	0.07%	Deficient	**0.006**	0.847 (P)	0.001 (D)	22.8	Conflicting
3	rs1049800	c.840T>C	Asp256Asp		0.22%	Unknown	0.953	NA	NA	0.085	Benign/Likely benign
3	rs17580	c.863A>T	Glu264Val	S	5.74%	Deficient	**1.15E-34**	0.998 (D)	0.0 (D)	23.7	Pathogenic
3	New2	c.894G>C	Lys274Asn		0.04%	Unknown	0.861	0.003 (B)	0.167 (T)	8.926	NA
4	rs141620200	c.922G>T	Ala284Ser		0.40%	Unknown	0.272	0.009 (B)	0.065 (T)	16.02	Conflicting
4	rs139964603	c.976G>A	Val302Ile		0.04%	Unknown	0.567	0.002 (B)	0.602 (T)	0.004	Conflicting
4	New3	c.1024C>T	Leu318Phe		0.04%	Unknown	0.252	0.776 (P)	0.053 (T)	24.9	NA
4	rs201788603	c.1061C>T	Ser330Phe	S_{Munich}	0.04%	Conflicting	0.764	1.0 (D)	0.0 (D)	27.3	Uncertain significance
5	rs9630	c.1068C>T	Ala332Ala		0.07%	Unknown	0.881	NA	NA	0.068	Benign
5	rs201318727	c.1090A>G	Ile340Val		0.04%	Unknown	0.726	0.001 (B)	0.244 (T)	4.739	NA
5	rs28929474	c.1096G>A	Glu342Lys	Z	1.69%	Deficient	**6.23E-51**	1.0 (D)	0.006 (D)	20.2	Pathogenic
5	rs61761869	c.1177C>T	Pro369Ser	$M_{Wurzburg}$	0.04%	Deficient	**0.033**	1.0 (D)	0.003 (D)	23.4	Conflicting
5	rs1303	c.1200A>C	Glu376Asp	M3 / M2	25.50%	Normal	**0.008**	0.0 (B)	1.0 (T)	0.001	Benign/Likely benign

MAF, minor allele frequency. The grey background in entire rows indicates known deficient alleles. Three additional variants considered deleterious by all pathogenicity scores are also in grey background. Note that rs112030253 in exon 2 is indicated twice as it is a multi-allelic polymorphism and three alleles are observed in CanCOLD.

*Genetic association with AAT serum levels. [†]D, probably damaging; P, possibly damaging; B, benign. [‡]D, damaging; T, tolerated.

Table 2. Allele distribution sorted by allele frequency in the overall CanCOLD cohort and by recruitment sites sorted by geographic location from west to east

Allele	CanCOLD	Vancouver	Calgary	Saskatoon	Toronto	Kingston	Ottawa	Montreal	Quebec	Halifax
M1	1246 (45.84)	391 (48.39)	119 (50)	83 (45.6)	56 (43.75)	84 (42.86)	95 (46.57)	255 (45.05)	63 (38.41)	100 (43.1)
M1(ala)	475 (17.48)	134 (16.58)	43 (18.07)	34 (18.68)	19 (14.84)	31 (15.82)	37 (18.14)	90 (15.9)	30 (18.29)	57 (24.57)
M2	396 (14.57)	114 (14.11)	36 (15.13)	22 (12.09)	26 (20.31)	33 (16.84)	25 (12.25)	88 (15.55)	28 (17.07)	24 (10.34)
M3	246 (9.05)	79 (9.78)	18 (7.56)	14 (7.69)	11 (8.59)	18 (9.18)	19 (9.31)	50 (8.83)	15 (9.15)	22 (9.48)
S	154 (5.67)	29 (3.59)	12 (5.04)	7 (3.85)	5 (3.91)	14 (7.14)	10 (4.9)	53 (9.36)	16 (9.76)	8 (3.45)
M3(Riedenburg)	48 (1.77)	12 (1.49)	2 (0.84)	4 (2.2)	5 (3.91)	4 (2.04)	5 (2.45)	6 (1.06)	5 (3.05)	5 (2.16)
Z	46 (1.69)	20 (2.48)	1 (0.42)	5 (2.75)	1 (0.78)	3 (1.53)	3 (1.47)	6 (1.06)	1 (0.61)	6 (2.59)
M4	36 (1.32)	6 (0.74)	3 (1.26)	7 (3.85)	2 (1.56)	6 (3.06)	2 (0.98)	4 (0.71)	1 (0.61)	5 (2.16)
M1(ala).Ala284Ser	11 (0.4)	3 (0.37)	0 (0)	1 (0.55)	0 (0)	1 (0.51)	4 (1.96)	1 (0.18)	0 (0)	1 (0.43)
M2(Obernburg)	8 (0.29)	1 (0.12)	0 (0)	1 (0.55)	0 (0)	0 (0)	0 (0)	4 (0.71)	2 (1.22)	0 (0)
F	6 (0.22)	3 (0.37)	2 (0.84)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
I	6 (0.22)	2 (0.25)	1 (0.42)	0 (0)	0 (0)	0 (0)	1 (0.49)	1 (0.18)	1 (0.61)	0 (0)
M1.Asp256Asp	5 (0.18)	3 (0.37)	0 (0)	0 (0)	1 (0.78)	0 (0)	1 (0.49)	0 (0)	0 (0)	0 (0)
M1(ala).Leu-10Leu	3 (0.11)	1 (0.12)	0 (0)	0 (0)	1 (0.78)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
V	3 (0.11)	0 (0)	0 (0)	1 (0.55)	0 (0)	1 (0.51)	0 (0)	1 (0.18)	0 (0)	0 (0)
M1(ala).Leu-2Gln	2 (0.07)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.86)
M1.Phe33Phe	2 (0.07)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.49)	0 (0)	1 (0.61)	0 (0)
M6(Passau)	2 (0.07)	1 (0.12)	0 (0)	0 (0)	0 (0)	1 (0.51)	0 (0)	0 (0)	0 (0)	0 (0)
P(Lowell)	2 (0.07)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.49)	0 (0)	0 (0)	0 (0)
M(Procida)	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M(Wurzburg)	1 (0.04)	0 (0)	0 (0)	0 (0)	1 (0.78)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M1(ala).Ala332Ala	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
M1(ala).Asp256Asp	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
M1(ala).Lys274Asn	1 (0.04)	0 (0)	0 (0)	1 (0.55)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M1(ala).Pro88Leu	1 (0.04)	0 (0)	1 (0.42)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M1(Cadiz)	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.43)
M1.Ala332Ala	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.61)	0 (0)
M1.Ala58Thr	1 (0.04)	0 (0)	0 (0)	1 (0.55)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M1.Leu318Phe	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
M1.Lys222Lys	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M1.Thr113Thr	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
M1.Val302Ile	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M3.Gly167Gly	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M3.Pro255Thr	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
M3.Ser-19Ser	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
R	1 (0.04)	0 (0)	0 (0)	1 (0.55)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
R.Ile340Val	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
S(Donosti)	1 (0.04)	1 (0.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
S(Munich)	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.43)
S.Arg101Cys	1 (0.04)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.18)	0 (0)	0 (0)
Total	2718 (100)	808 (100)	238 (100)	182 (100)	128 (100)	196 (100)	204 (100)	566 (100)	164 (100)	232 (100)
Normal allele	2498 (91.91)	750 (92.82)	222 (93.28)	170 (93.41)	121 (94.53)	179 (91.33)	189 (92.65)	504 (89.05)	146 (89.02)	217 (93.53)
Deficient allele	220 (8.09)	58 (7.18)	16 (6.72)	12 (6.59)	7 (5.47)	17 (8.67)	15 (7.35)	62 (10.95)	18 (10.98)	15 (6.47)

Values are allele counts with percentage by site in parentheses. The grey background indicates known deficient alleles or considered deleterious by all pathogenicity scores.

Table 3. Clinical characteristics by *SERPINA1* genotyping groups

Characteristics	Group 1: no deficiency n=1,149	Group 2: mild deficiency n=147	Group 3: intermediate deficiency n=59	Group 4: severe deficiency n=4	P value*	P value adj. age, sex, and study sites	P value for interaction with smoking†
Age (years; mean ± SD)	66.8 ± 9.7	64.9 ± 9.6	68.8 ± 8.6	68.3 ± 10.6	0.057	-	-
AAT levels (g/L)	1.31 ± 0.17 [12][a]	1.10 ± 0.14 [1][b]	0.96 ± 0.20 [2][c]	0.50 ± 0.21[d]	**<0.001**	**<0.001**	0.445
Sex – no. of patients (%)					0.553		-
Male	647 (56.3)	85 (57.8)	36 (61.0)	1 (25.0)			
Female	502 (43.7)	62 (42.2)	23 (39.0)	3 (75.0)			
BMI (kg/m^2)	27.6 ± 5.0 [1]	27.9 ± 5.3	27.0 ± 4.3	25.7 ± 4.8	0.650	0.881	0.876
Smoking status – no. of patients (%)					0.692	0.801	-
Never	401 (34.9)	55 (37.4)	15 (25.4)	2 (50.0)			
Former	561 (48.8)	70 (47.6)	35 (59.3)	2 (50.0)			
Current	187 (16.3)	22 (15.0)	9 (15.3)	0 (0.0)			
Pack-years (ever-smokers)	16.7 ± 22.5 [19]	16.9 ± 21.9 [1]	16.2 ± 19.8 [1]	11.7 ± 22.2	0.954	0.982	-
Post-bronchodilator FEV1, L	2.55 ± 0.81	2.62 ± 0.88	2.58 ± 0.82	2.18 ± 0.62	0.693	0.921	0.424
Post-bronchodilator FEV1 % predicted	91.5 ± 20.4	90.8 ± 19.7	91.9 ± 21.5	90.1 ± 27.8	0.883	0.992	0.33
Post-FEV1/FVC	69.5 ± 10.4	69.4 ± 9.8	68.9 ± 12.4	59.7 ± 13.0	0.406	0.271	0.464
FEV1 reversibility, %	5.4 ± 7.9	6.7 ± 8.9	5.1 ± 10.9	7.9 ± 12.9	0.284	0.346	0.553
FEF25-75	1.77 ± 1.04	1.79 ± 1.04	1.79 ± 1.00	1.05 ± 0.68	0.492	0.519	0.662
DLCO, % predicted	106.4 ± 24.3 [56][a]	109.9 ± 26.5 [11][a]	102.5 ± 25.3 [6][a]	69.4 ± 19.1[b]	**0.005**	**0.008**	0.289
RV, % predicted	120.1 ± 34.8 [61]	125.1 ± 34.9 [10]	117.1 ± 32.7 [3]	145.9 ± 33.1	0.099	0.12	0.636
FRC, % predicted	110.3 ± 25.6 [54]	113.6 ± 25.2 [10]	108.6 ± 25.6 [3]	136.9 ± 19.7	0.052	0.067	0.892
TLC, % predicted	118.1 ± 17.4 [52]	120.2 ± 18.3 [11]	116.2 ± 15.7 [3]	133.8 ± 11.9	0.127	0.097	0.942
LAA-950 (%)	4.30 ± 4.52 [135][a]	4.11 ± 4.23 [25][a]	4.22 ± 4.92 [2][a]	19.89 ± 12.18[b]	**0.014**	**<0.001**	0.169
PD15	-917.2 ± 21.5 [135][a]	-916.5 ± 19.8 [25][a]	-918.7 ± 18.6 [2][a]	-954.5 ± 15.0[b]	**0.015**	**0.003**	0.939
Pi10_all	3.96 ± 0.16 [135]	3.97 ± 0.14 [25]	3.99 ± 0.17 [2]	4.00 ± 0.14	0.34	0.283	0.817
Pi10_leq20	3.87 ± 0.11 [135]	3.88 ± 0.11 [25]	3.90 ± 0.12 [2]	3.90 ± 0.08	0.478	0.371	0.714
MRC scales (Level 1-5)	1.44 ± 0.65 [51]	1.40 ± 0.67 [11]	1.49 ± 0.81 [4]	1.75 ± 0.50	0.343	0.761	0.634
MRC scales ≥ 3, n (%)	58 (5.3) [51]	8 (5.9) [11]	6 (10.9) [4]	0 (0.0)	0.335	0.421	0.996
CAT score	6.8 ± 5.9 [25]	7.9 ± 6.7 [3]	6.5 ± 8.1	5.8 ± 2.9	0.078	0.131	0.42
SGRQ-Total score	13.3 ± 14.4 [255]	15.5 ± 17.0 [26]	15.2 ± 19.0 [15]	20.0 ± 4.4 [1]	0.332	0.299	0.302
Self-reported comorbidities, n (%)							
Hypertension	402 (35.0)	51 (34.7)	19 (32.2)	2 (50.0)	0.88	0.699	0.624
CVD (any CVD excluding Hypertension)	325 (28.3)	40 (27.2)	21 (35.6)	1 (25.0)	0.618	0.854	0.345
Diabetes	123 (10.7)	13 (8.8)	8 (13.6)	0 (0.0)	0.724	0.903	**0.021**
Asthma	267 (23.2)	32 (21.8)	16 (27.1)	1 (25.0)	0.813	0.725	0.57
Respiratory medications, n (%)							
SAMA/SABA	35 (3.0)	6 (4.1)	5 (8.5)	0 (0.0)	0.128	0.163	1

LABA± SAMA/SABA	3 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	1.000	1.000	1
LAMA± SAMA/SABA	10 (0.9)	3 (2.0)	1 (1.7)	0 (0.0)	0.248	0.567	1
LAMA+LABA± SAMA/SABA	71 (6.2)	11 (7.5)	2 (3.4)	0 (0.0)	0.690	0.826	0.485
ICS± SAMA/SABA	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1.000	1.000	1
LABA+ICS± SAMA/SABA	99 (8.6)	10 (6.8)	1 (1.7)	1 (25.0)	0.098	0.232	0.442
LAMA+ICS± SAMA/SABA	3 (0.3)	1 (0.7)	0 (0.0)	0 (0.0)	0.490	0.819	1
LAMA+LABA+ICS± SAMA/SABA	32 (2.8)	5 (3.4)	3 (5.1)	0 (0.0)	0.469	0.799	1
Any above medications	255 (22.2)	36 (24.5)	12 (20.3)	1 (25.0)	0.855	0.936	0.995

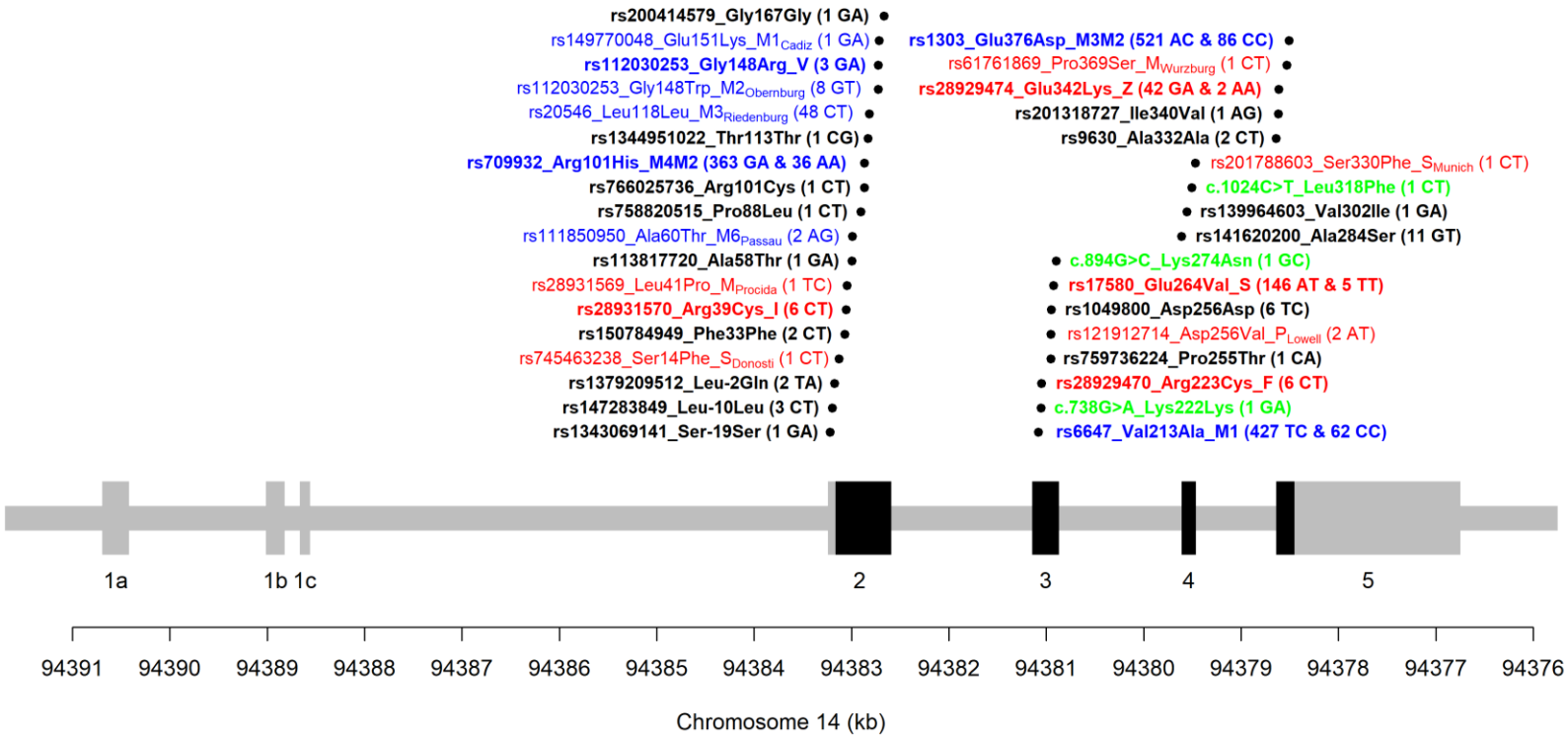
*Post-hoc Tukey multiple pairwise-comparison tests were used for statistically significant ANOVA. Means with different letters indicate groups that significantly differed.

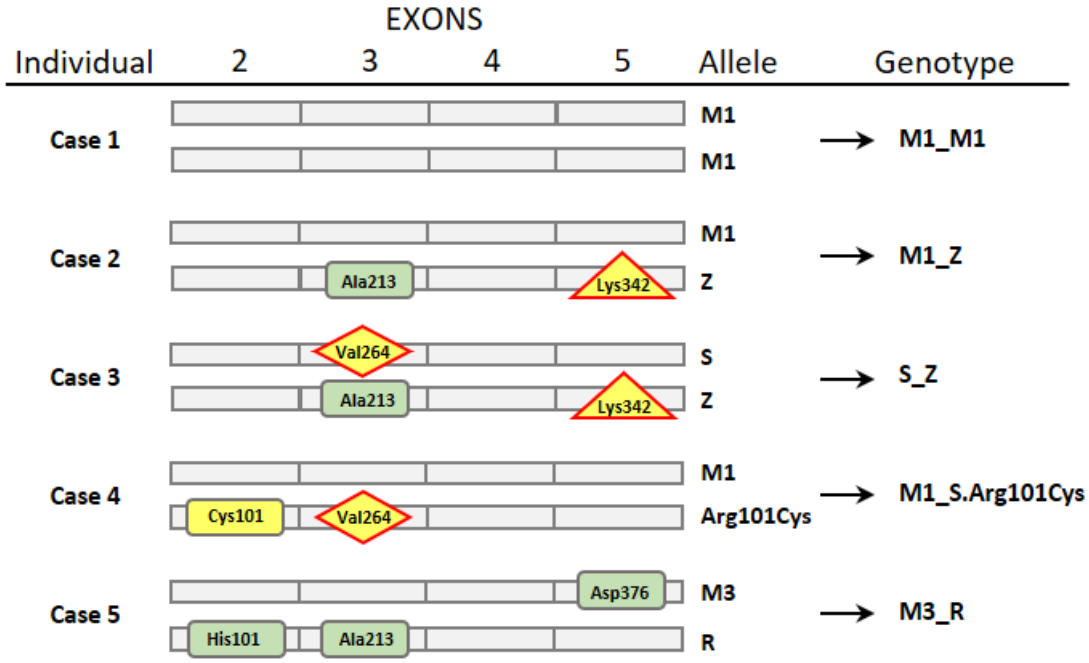
†P value for the interaction term between *SERPINA1* genotyping groups and smoking (ever vs. never smokers), adjusted for age, sex, and study sites.

The number of missing values is indicated in square brackets. For lung function measurements, the maximum values among all the trials are indicated.

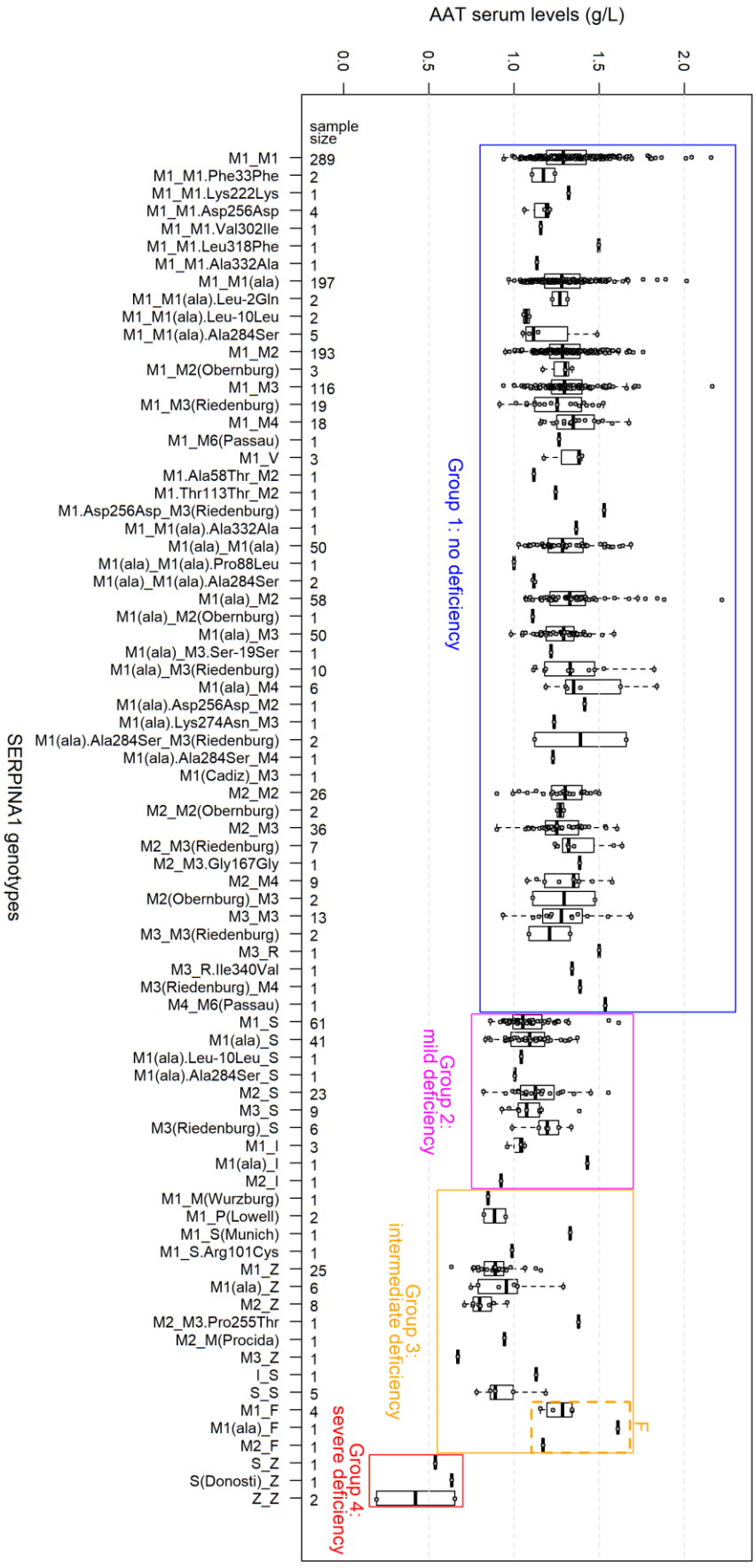
Group 1 includes all individuals with two normal alleles (n=1,149). Group 2 are individuals with one normal allele and one S or I allele (n=147, MS and MI). Group 3 consists of individuals that are heterozygotes (MZ, MF, $MM_{Procida}$, $MM_{Wurzburg}$, MP_{Lowell}, MS_{Munich}, MS.Arg101Cys, MM3.Pro255Thr), homozygotes for the S allele (SS) and compound heterozygotes for deficient alleles (IS) (n=59). Group 4 are AATD individuals (n=4, ZZ, SZ, and $S_{Donosti}Z$).

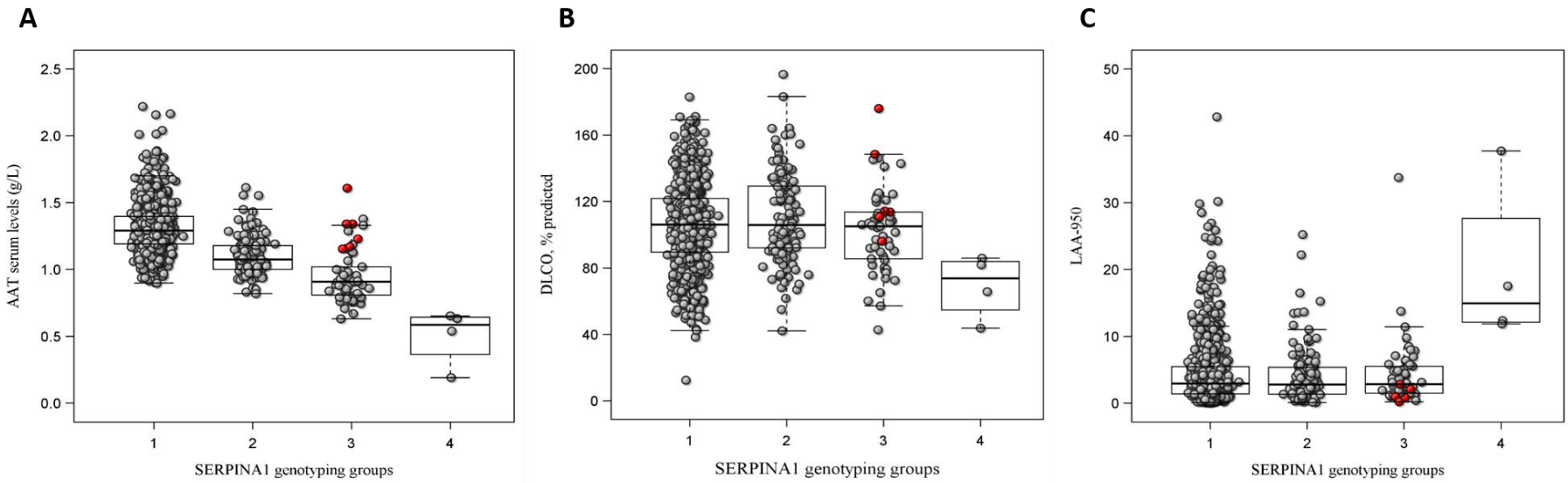
AAT: alpha-1 antitrypsin; BMI, body mass index; CAT, COPD Assessment Test; CVD, cardiovascular diseases; DLCO: diffusion capacity to carbon monoxide; FEF25-75: forced expiratory flow at 25–75%; FEV1: forced expiratory volume in 1 second; FRC: functional residual capacity; FVC: forced vital capacity; ICS, inhaled corticosteroid; LAA-950: percentage of lung voxels below -950 Hounsfield units; LABA, long-acting β-agonist; LAMA, long-acting muscarinic antagonist; MRC, modified British Medical Research Council Questionnaire; PD15: 15th percentile of the CT attenuation histogram; Pi10_all: airway wall thickness of an airway with internal perimeter of 10mm generated using all CT airways; Pi10_leq20: airway wall thickness of a theoretical airway with internal perimeter of 10mm generated using all CT airways with an internal perimeter <=20mm; RV: residual volume; SABA, short-acting β-agonist; SAMA, short-acting muscarinic antagonist; SGRQ, St. George's Respiratory Questionnaire for COPD; TLC: total lung capacity.


rs200414579_Gly167Gly (1 GA)
rs149770048_Glu151Lys_M1Cadiz (1 GA)
rs112030253_Gly148Arg_V (3 GA)
rs112030253_Gly148Trp_M2Oberndurg (8 GT)
rs20546_Leu118Leu_M3Riedenburg (48 CT)
rs1344951022_Thr113Thr (1 CG)
rs709932_Arg101His_M4M2 (363 GA & 36 AA)
rs766025736_Arg101Cys (1 CT)
rs758820515_Pro88Leu (1 CT)
rs111850950_Ala60Thr_M6Passau (2 AG)
rs113817720_Ala58Thr (1 GA)
rs28931569_Leu41Pro_MProcida (1 TC)
rs28931570_Arg39Cys_I (6 CT)
rs150784949_Phe33Phe (2 CT)
rs745463238_Ser14Phe_SDonosti (1 CT)
rs1379209512_Leu-2Gln (2 TA)
rs147283849_Leu-10Leu (3 CT)
rs1343069141_Ser-19Ser (1 GA)
rs1303_Glu376Asp_M3M2 (521 AC & 86 CC)
rs61761869_Pro369Ser_MWurzburg (1 CT)
rs28929474_Glu342Lys_Z (42 GA & 2 AA)
rs201318727_Ile340Val (1 AG)
rs9630_Ala332Ala (2 CT)
rs201788603_Ser330Phe_SMunich (1 CT)
c.1024C>T_Leu318Phe (1 CT)
rs139964603_Val302Ile (1 GA)
rs141620200_Ala284Ser (11 GT)
c.894G>C_Lys274Asn (1 GC)
rs17580_Glu264Val_S (146 AT & 5 TT)
rs1049800_Asp256Asp (6 TC)
rs121912714_Asp256Val_PLowell (2 AT)
rs759736224_Pro255Thr (1 CA)
rs28929470_Arg223Cys_F (6 CT)
c.738G>A_Lys222Lys (1 GA)
rs6647_Val213Ala_M1 (427 TC & 62 CC)
1a
1b 1c
2
3
4
5
94391
94390
94389
94388
94387
94386
94385
94384
94383
94382
94381
94380
94379
94378
94377
94376
Chromosome 14 (kb)


EXONS
Individual
2
3
4
5
Allele
Genotype
Case 1
M1
M1
M1_M1
Case 2
M1
Ala213
Lys342
Z
M1_Z
Case 3
Val264
S
Ala213
Lys342
Z
S_Z
Case 4
M1
Cys101
Val264
Arg101Cys
M1_S.Arg101Cys
Case 5
Asp376
M3
His101
Ala213
R
M3_R

Allele	R*	M3.Pro255Thr	S.Arg101Cys	V	$M1_{Cadiz}$	S_{Munich}	P_{Lowell}	I	F	$M_{Wurzburg}$	$S_{Donosti}$	Z	S	$M3_{Riedenburg}$	$M2_{Obernburg}$	$M_{Procida}$	$M6_{Passau}$	M4	M3	M2	M1	$M1(Ala^{213})$	
$M1(Ala^{213})$	0	0	0	0	0	0	0	1	1	0	0	6	43	12	1	0	0	7	52	59	207	53	
M1	0	0	1	3	0	1	2	3	4	1	0	25	61	20	3	0	1	18	116	195	299		
M2	0	1	0	0	0	0	0	1	1	0	0	8	23	7	2	1	0	9	37	26			
M3	2	0	0	0	1	0	0	0	0	0	0	1	9	2	2	0	0	0	13				
M4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0					
$M6_{Passau}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
$M_{Procida}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
$M2_{Obernburg}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
$M3_{Riedenburg}$	0	0	0	0	0	0	0	0	0	0	0	0	6	0									
S	0	0	0	0	0	0	0	1	0	0	0	1	5										
Z	0	0	0	0	0	0	0	0	0	0	1	2											
$S_{Donosti}$	0	0	0	0	0	0	0	0	0	0	0												
$M_{Wurzburg}$	0	0	0	0	0	0	0	0	0	0													
F	0	0	0	0	0	0	0	0	0														
I	0	0	0	0	0	0	0	0															
P_{Lowell}	0	0	0	0	0	0	0																
S_{Munich}	0	0	0	0	0	0																	
$M1_{Cadiz}$	0	0	0	0	0																		
V	0	0	0	0																			
S.Arg101Cys	0	0	0																				
M3.Pro255Thr	0	0																					
R*	0																						1359


AAT serum levels (g/L)
0.0
0.5
1.0
1.5
2.0
SERPINA1 genotypes
sample size
M1_M1 289
M1_M1.Phe33Phe 2
M1_M1.Lys222Lys 1
M1_M1.Asp256Asp 4
M1_M1.Val302Ile 1
M1_M1.Leu318Phe 1
M1_M1.Ala332Ala 1
M1_M1(ala) 197
M1_M1(ala).Leu-2Gln 2
M1_M1(ala).Leu-10Leu 2
M1_M1(ala).Ala284Ser 5
M1_M2 193
M1_M2(Obernburg) 3
M1_M3 116
M1_M3(Riedenburg) 19
M1_M4 18
M1_M6(Passau) 1
M1_V 3
M1.Ala58Thr_M2 1
M1.Thr113Thr_M2 1
M1.Asp256Asp_M3(Riedenburg) 1
M1_M1(ala).Ala332Ala 1
M1(ala)_M1(ala) 50
M1(ala)_M1(ala).Pro88Leu 1
M1(ala)_M1(ala).Ala284Ser 2
M1(ala)_M2 58
M1(ala)_M2(Obernburg) 1
M1(ala)_M3 50
M1(ala)_M3.Ser-19Ser 1
M1(ala)_M3(Riedenburg) 10
M1(ala)_M4 6
M1(ala).Asp256Asp_M2 1
M1(ala).Lys274Asn_M3 1
M1(ala).Ala284Ser_M3(Riedenburg) 2
M1(ala).Ala284Ser_M4 1
M1(Cadiz)_M3 1
M2_M2 26
M2_M2(Obernburg) 2
M2_M3 36
M2_M3(Riedenburg) 7
M2_M3.Gly167Gly 1
M2_M4 9
M2(Obernburg)_M3 2
M3_M3 13
M3_M3(Riedenburg) 2
M3_R 1
M3_R.Ile340Val 1
M3(Riedenburg)_M4 1
M4_M6(Passau) 1
M1_S 61
M1(ala)_S 41
M1(ala).Leu-10Leu_S 1
M1(ala).Ala284Ser_S 1
M2_S 23
M3_S 9
M3(Riedenburg)_S 6
M1_I 3
M1(ala)_I 1
M2_I 1
M1_M(Wurzburg) 1
M1_P(Lowell) 2
M1_S(Munich) 1
M1_S.Arg101Cys 1
M1_Z 25
M1(ala)_Z 6
M2_Z 8
M2_M3.Pro255Thr 1
M2_M(Procida) 1
M3_Z 1
I_S 1
S_S 5
M1_F 4
M1(ala)_F 1
M2_F 1
S_Z 1
S(Donosti)_Z 1
Z_Z 2
Group 1: no deficiency
Group 2: mild deficiency
Group 3: intermediate deficiency
F
Group 4: severe deficiency


A
AAT serum levels (g/L)
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
SERPINA1 genotyping groups
B
DLCO, % predicted
200
160
120
80
40
0
1
2
3
4
SERPINA1 genotyping groups
C
LAA-950
50
40
30
20
10
0
1
2
3
4
SERPINA1 genotyping groups

Supplementary material

Granularity of *SERPINA1* alleles by DNA sequencing in CanCOLD

Nisha Gupta, MD; Nathalie Gaudreault, BSc; Sébastien Thériault, MD; Pei Zhi Li, MSc; Cyndi Henry, MSc; Miranda Kirby, PhD; François Maltais, MD; Wan Tan, MD; Jean Bourbeau, MD; Yohan Bossé, PhD; for the Canadian Respiratory Research Network, for the CanCOLD study group

Methods

Pulmonary function tests

Spirometry and plethysmography were performed according to the American Thoracic Society guidelines [1-3]. For spirometry, FEV1, FVC, and FEV1/FVC were reported. Whole-body plethysmography was also performed for the measurement of functional residual capacity (FRC), residual volume (RV), total lung capacity (TLC), and diffusing capacity of the lung for carbon monoxide (DLCO).

CT image acquisition and analysis

CT images were acquired at multiple sites using various CT system models calibrated similarly with the participant supine at suspended full inspiration and full expiration from the apex to the base of the lung. The CT parameters for image acquisition in the participants investigated were as follows: 100 kVp, 50 mAs, 0.5 s gantry rotation, pitch of 1.375, 1.00–1.25 mm slice thickness, and an intermediate reconstruction kernel (GE: Standard; Siemens: b35; Philips: B) was used for quantitative analysis. All CT image analysis was performed by VIDA Diagnostics, Inc. (Coralville, IA, USA). CT emphysema was measured in the full-inspiration CT images using the percentage of low attenuation areas below -950 Hounsfield units (LAA-950) [4, 5].

DNA sequencing of *SERPINA1*

DNA was extracted from 200-400 µL of frozen whole blood using QIAamp® DNA Blood Mini kit (Qiagen). The DNA quality and concentration were assessed by the UV absorbance ratio 260 nm/280 nm and UV absorbance 260 nm, respectively. Every sample was diluted to a final concentration of 50 ng/µL. The DNA sequences of the coding regions (i.e. exons 2 to 5) of the *SERPINA1* gene were obtained by Sanger sequencing for all subjects. Primer sequences to evaluate the selected regions of *SERPINA1* are provided in **Table S4**. PCR was performed in a final volume of 25 µL containing 100

ng of genomic DNA, 1 U of HotStarTaq DNA polymerase (Qiagen), PCR buffer 1X, Q-Solution 1X, 160 μM of each dNTP and 0.2 μM of each primer. The PCR reaction was carried out on the GeneAmp® PCR system 9700 (Applied Biosystems) with the following cycling conditions: 15 minutes at 95°C, 35 PCR amplification cycles (15 seconds at 94°C, 30 seconds at 59°C (exons 2, 3 and 4) or 60°C (exon 5), and 60 seconds at 72°C), and 7 minutes at 72°C. The sequencing reaction was then performed using standard procedures and the product was run on the ABI 3730xl DNA Analyzer (Applied Biosystems). Sequencing files were assembled and analyzed using the EMBL-EBI Clustal Omega Multiple Alignment Tool (http://www.ebi.ac.uk/Tools/msa/clustalo).

Allele-specific PCR for allelic background determination

For identified variants with unknown allelic background, allele-specific PCR (AS-PCR) was performed in order to amplify and sequence each allele independently. For each variant, two forward AS-primers and one common reverse primer were designed based on the heterozygosity of a single polymorphism. Furthermore, the resulting amplicon was designed in order to include all polymorphisms identified in a single DNA sample of interest. For example, rs201318727 (Ile340Val), rs1303 (Glu376Asp) and rs709932 (Arg101His) were all found heterozygote in the same DNA sample. The design was based on rs1303 heterozygosity and the amplicon spanned 4,598 bp in order to include the other two polymorphisms. Each allele-specific amplicon was then sequenced using primers located near each polymorphism to be sequenced. **Table S5** shows the AS-PCR primers and the polymorphism used to design the AS-PCR. Primers were purchased from Integrated DNA Technologies, Iowa, USA. PCR cycling conditions are indicated in **Table S6**. DNA sequencing was performed as indicated above.

Results

Association of individual genetic variants with AAT serum levels

All the genetic variants identified were tested against the AAT serum levels. The most significant associated variant was rs28929474 causing the Z phenotype (**Table 1**). Mean AAT levels by genotype groups were 1.28±0.18 for GG (no Z allele), 0.87±0.15 for GA (heterozygote for the Z allele), and 0.42±0.32 for AA (homozygote for the Z allele). The second most significant associated variant was rs17580 causing the S phenotype. Mean AAT levels by genotype groups were 1.29±0.19 for AA, 1.10±0.15 for AT, and 0.94±0.16 for TT. The three common missense variants defining the M1 to M4 phenotypes were also associated with AAT levels. For rs6647 (Val213Ala), the mean AAT levels by genotype groups were 1.28±0.18 for TT, 1.25±0.22 for TC, and 1.23±0.25 for CC. For rs709932 (Arg101His), the mean AAT levels by genotype groups were 1.26±0.20 for GG, 1.30±0.19 for GA, and 1.29±0.16 for AA. For rs1303 (Glu376Asp), the mean AAT levels by genotype groups were 1.26±0.21 for AA, 1.29±0.18 for AC, and 1.29±0.16 for CC. Four additional missense variants were associated with lowered AAT levels including rs28931570 (Arg39Cys) causing the I phenotype, rs121912714 (Asp256Val) causing P_{Lowell}, rs61761869 (Pro369Ser) causing $M_{Wurzburg}$, and rs745463238 (Ser14Phe), but the latter occurs in only one individual with $ZS_{Donosti}$.

Deficient alleles across CanCOLD subgroups

We next evaluated the distribution of the 220 deficient alleles across the four CanCOLD subgroups, i.e. controls, at-risk, mild COPD, and moderate-severe COPD. Deficient alleles were found in 7.7% (47 out of 612) of controls, 7.3% (59 out of 808) of individuals at-risk of COPD, 8.9% (64 out of 718) of mild COPD patients, and 8.6% (50 out of 580) of moderate-severe COPD patients. A slight increase of deficient alleles was observed in the COPD subgroups, however this difference was not statistically significant (**Figure S1A**). Similarly, the distribution of the S and Z alleles across CanCOLD subgroups were not statistically different (**Figures S1B-C**).

Carriers of deficient alleles across CanCOLD subgroups

Finally, we evaluated the distribution of the 210 individuals carrying at least one deficient alleles across CanCOLD subgroups. There were 14.4% (44 out of 306) of carriers in controls, 14.1% (57 out of 404) in at-risk individuals, 17.0% (61 out of 359) in mild COPD patients, and 16.6% (48 out of 290) in moderate-severe COPD patients. A slight increase of carriers of deficient alleles was observed in the COPD subgroups, however the difference was not statistically significant (**Figure S2A**). Similarly, carriers of the S and Z alleles were not statistically different across CanCOLD subgroups (**Figures S2B-C**). When considering those with or without emphysema according to CT imaging, the percentages of AATD allele carriers was 15.2% (129 out of 848) for LAA-950 <5% and 15.5% (54 out of 349) for LAA-950 ≥5%.

References

1. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, Gustafsson P, Hankinson J, Jensen R, McKay R, Miller MR, Navajas D, Pedersen OF, Pellegrino R, Wanger J. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 2005: 26(4): 720-735.
2. Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, Casaburi R, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Hankinson J, Jensen R, Johnson D, Macintyre N, McKay R, Miller MR, Navajas D, Pellegrino R, Viegi G. Standardisation of the measurement of lung volumes. *Eur Respir J* 2005: 26(3): 511-522.
3. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J, Force AET. Standardisation of spirometry. *Eur Respir J* 2005: 26(2): 319-338.
4. Gevenois PA, Zanen J, de Maertelaer V, De Vuyst P, Dumortier P, Yernault JC. Macroscopic assessment of pulmonary emphysema by image analysis. *J Clin Pathol* 1995: 48(4): 318-322.
5. Muller NL, Staples CA, Miller RR, Abboud RT. "Density mask". An objective method to quantitate emphysema using computed tomography. *Chest* 1988: 94(4): 782-787.

Table S1. Clinical characteristics of the CanCOLD participants

Characteristics	All	Healthy controls	At-risk of COPD	Mild COPD	Moderate-severe COPD
Number of patients	1,359	306	404	359	290
Age (years; mean ± SD)	66.6 ± 9.7	66.7 ± 9.4	65.5 ± 9.4	67.9 ± 9.6	66.6 ± 10.2
AAT levels (g/L)	1.27 ± 0.20 [15]	1.24 ± 0.18 [3]	1.28 ± 0.20 [7]	1.25 ± 0.19 [2]	1.31 ± 0.21 [3]
Sex – no. of patients (%)					
Male	769 (56.6)	144 (47.1)	229 (56.7)	237 (66.0)	159 (54.8)
Female	590 (43.4)	162 (52.9)	175 (43.3)	122 (34.0)	131 (45.2)
BMI (kg/m^2)	27.6 ± 5.0 [1]	27.3 ± 4.9	28.0 ± 5.2 [1]	27.0 ± 4.5	27.8 ± 5.4
Smoking status – no. of patients (%)					
Never	473 (34.8)	306 (100.0)	0 (0.0)	114 (31.8)	53 (18.3)
Former	668 (49.2)	0 (0.0)	312 (77.2)	196 (54.6)	160 (55.2)
Current	218 (16.0)	0 (0.0)	92 (22.8)	49 (13.6)	77 (26.6)
Pack-years (ever-smokers)	16.7 ± 22.3 [21]	0.0 ± 0.0	18.8 ± 18.7 [14]	18.3 ± 23.1 [2]	29.9 ± 26.8 [5]
Post-bronchodilator FEV1, L	2.6 ± 0.8	2.7 ± 0.8	2.8 ± 0.8	2.7 ± 0.7	1.8 ± 0.6
Post-bronchodilator FEV1 % predicted	91.4 ± 20.3	101.6 ± 17.1	99.2 ± 16.6	95.6 ± 11.9	64.8 ± 12.2
Post-FEV1/FVC	69.4 ± 10.4	77.5 ± 4.7	76.8 ± 4.5	64.5 ± 4.6	56.7 ± 9.6
FEV1 reversibility, %	5.5 ± 8.2	3.5 ± 5.7	3.3 ± 5.7	6.5 ± 6.2	9.6 ± 12.7
FEF25-75	1.8 ± 1.0	2.4 ± 0.9	2.4 ± 1.0	1.3 ± 0.5	0.7 ± 0.4
DLCO, % predicted	106.8 ± 26.7 [72]	112.9 ± 24.0 [12]	110.8 ± 29.6 [27]	108.8 ± 23.0 [12]	91.8 ± 24.5 [21]
RV, % predicted	120.5 ± 34.7 [74]	109.0 ± 27.9 [15]	110.4 ± 26.9 [32]	122.8 ± 32.1 [14]	143.4 ± 41.6 [13]
FRC, % predicted	110.7 ± 25.6 [67]	105.6 ± 24.7 [14]	103.6 ± 21.1 [27]	114.8 ± 22.9 [14]	120.5 ± 30.8 [12]
TLC, % predicted	118.2 ± 17.5 [66]	115.2 ± 18.8 [13]	115.1 ± 15.1 [27]	125.2 ± 16.4 [13]	117.0 ± 17.9 [13]
LAA-950	4.3 ± 4.6 [182]	3.0 ± 3.1 [60]	3.0 ± 3.2 [46]	5.6 ± 5.0 [31]	6.0 ± 6.1 [45]
PD15	-917.4 ± 21.3 [162]	-911.2 ± 21.7 [52]	-911.3 ± 20.5 [40]	-924.9 ± 17.8 [27]	-922.5 ± 21.8 [43]
Pi10_all	4.0 ± 0.2 [162]	3.9 ± 0.2 [52]	4.0 ± 0.1 [40]	4.0 ± 0.2 [27]	4.0 ± 0.2 [43]
Pi10_leq20	3.9 ± 0.1 [162]	3.8 ± 0.1 [52]	3.9 ± 0.1 [40]	3.9 ± 0.1 [27]	3.9 ± 0.1 [43]
MRC scales (Level 1-5)	1.4 ± 0.6 [137]	1.3 ± 0.6 [18]	1.3 ± 0.6 [46]	1.3 ± 0.5 [29]	1.8 ± 0.8 [44]
MRC scales ≥ 3, n (%)	59 (4.8) [137]	6 (2.1) [18]	10 (2.8) [46]	7 (2.1) [29]	36 (14.6) [44]
CAT score	6.9 ± 6.1 [28]	5.5 ± 4.3 [10]	6.1 ± 5.7 [14]	5.6 ± 4.9 [2]	10.8 ± 7.7 [2]
SGRQ-Total score	13.0 ± 14.5 [297]	8.1 ± 9.6 [242]	9.5 ± 11.9 [42]	9.4 ± 10.3 [12]	22.7 ± 17.6 [1]
Self-reported comorbidities, n (%)					
Hypertension	474 (34.9)	107 (35.0)	129 (31.9)	114 (31.8)	124 (42.8)
CVD (any CVD excluding Hypertension)	387 (28.5)	86 (28.1)	97 (24.0)	92 (25.6)	112 (38.6)
Diabetes	144 (10.6)	29 (9.5)	48 (11.9)	29 (8.1)	39 (13.1)
Asthma	316 (23.3)	46 (15.0)	67 (16.6)	84 (23.4)	119 (41.0)
Respiratory medications, n (%)					
SAMA/SABA	46 (3.4)	4 (1.3)	5 (1.2)	17 (4.7)	20 (6.9)

LABA± SAMA/SABA	3 (0.2)	1 (0.3)	0 (0.0)	1 (0.3)	1 (0.3)
LAMA± SAMA/SABA	14 (1.0)	0 (0.0)	2 (0.5)	1 (0.3)	11 (3.8)
LAMA+LABA± SAMA/SABA	84 (6.2)	12 (3.9)	20 (5.0)	26 (7.2)	26 (9.0)
ICS± SAMA/SABA	2 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.7)
LABA+ICS± SAMA/SABA	111 (8.2)	11 (3.6)	20 (5.0)	22 (6.1)	58 (20.0)
LAMA+ICS± SAMA/SABA	4 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	4 (1.4)
LAMA+LABA+ICS± SAMA/SABA	40 (2.9)	0 (0.0)	1 (0.2)	4 (1.1)	35 (12.1)
Any above medications	304 (22.4)	28 (9.2)	48 (11.9)	71 (19.8)	157 (54.3)

The number of missing values is indicated in square brackets. For lung function measurements, the maximum values among all the trials are indicated.

AAT: alpha-1 antitrypsin; BMI, body mass index; CAT, COPD Assessment Test; CVD, cardiovascular diseases; DLCO: diffusion capacity to carbon monoxide; FEF25-75: forced expiratory flow at 25–75%; FEV1: forced expiratory volume in 1 second; FRC: functional residual capacity; FVC: forced vital capacity; ICS, inhaled corticosteroid; LAA-950: percentage of lung voxels below -950 Hounsfield units; LABA, long-acting β-agonist; LAMA, long-acting muscarinic antagonist; MRC, modified British Medical Research Council Questionnaire; PD15: 15th percentile of the CT attenuation histogram; Pi10_all: airway wall thickness of an airway with internal perimeter of 10mm generated using all CT airways; Pi10_leq20: airway wall thickness of a theoretical airway with internal perimeter of 10mm generated using all CT airways with an internal perimeter <=20mm; RV: residual volume; SABA, short-acting β-agonist; SAMA, short-acting muscarinic antagonist; SGRQ, St. George's Respiratory Questionnaire for COPD; TLC: total lung capacity.

Table S2. Clinical characteristics by *SERPINA1* genotyping groups in ever- and never-smokers

Characteristics	Group 1: no deficiency	Group 2: mild deficiency	Group 3: intermediate deficiency	Group 4: severe deficiency	P value*	P value adj. age, sex, and study sites
Ever-smokers	**n=748**	**n=92**	**n=44**	**n=2**		
Age (years; mean ± SD)	66.6 ± 9.7	65.1 ± 9.4	68.6 ± 8.2	61.5 ± 4.9	0.208	-
AAT levels (g/L)	1.33 ± 0.17 [10][a]	1.12 ± 0.14[b]	0.98 ± 0.21 [1][c]	0.41 ± 0.31[d]	**<0.001**	**<0.001**
Sex – no. of patients (%)					0.306	
Male	441 (59.0)	59 (64.1)	27 (61.4)	0 (0.0)		
Female	307 (41.0)	33 (35.9)	17 (38.6)	2 (100.0)		
BMI (kg/m^2)	27.9 ± 5.1 [1]	28.4 ± 5.6	27.1 ± 4.6	28.3 ± 6.5	0.748	0.834
Smoking status – no. of patients (%)					0.909	0.96
Former	561 (75.0)	70 (76.1)	35 (79.5)	2 (100.0)		
Current	187 (25.0)	22 (23.9)	9 (20.5)	0 (0.0)		
Pack-years (ever-smokers)	26.0 ± 23.4 [19]	27.1 ± 22.2 [1]	21.9 ± 20.1 [1]	23.5 ± 30.5	0.621	0.757
Post-bronchodilator FEV1, L	2.5 ± 0.8	2.6 ± 0.8	2.5 ± 0.8	1.8 ± 0.4	0.463	0.854
Post-bronchodilator FEV1 % predicted	88.6 ± 20.9	88.5 ± 20.1	87.8 ± 22.0	72.4 ± 11.4	0.613	0.768
Post-FEV1/FVC	68.0 ± 10.8	67.6 ± 10.1	67.2 ± 13.2	52.2 ± 8.4	0.259	0.138
FEV1 reversibility, %	5.8 ± 8.2	7.1 ± 9.5	5.7 ± 12.3	14.3 ± 18.2	0.544	0.325
FEF25-75	1.7 ± 1.0	1.7 ± 1.0	1.7 ± 1.0	0.5 ± 0.2	0.302	0.387
DLCO, % predicted	103.9 ± 24.7 [42]	104.0 ± 22.7 [7]	100.7 ± 26.2 [4]	73.8 ± 11.4	0.308	0.329
RV, % predicted	123.4 ± 36.2 [42]	129.4 ± 37.0 [7]	122.5 ± 35.6 [3]	164.3 ± 41.9	0.209	0.128
FRC, % predicted	111.6 ± 26.0 [36]	114.6 ± 25.4 [7]	111.1 ± 28.3 [3]	145.3 ± 29.6	0.241	0.125
TLC, % predicted	118.7 ± 17.4 [35]	120.2 ± 15.5 [8]	116.9 ± 17.2 [3]	137.5 ± 19.1	0.354	0.248
LAA-950 (%)	4.6 ± 4.9 [79][a]	4.3 ± 4.2 [15][a]	4.6 ± 5.6 [2][a]	14.7 ± 4.0[b]	**0.034**	**0.010**
PD15	-918.1 ± 21.8 [79]	-917.3 ± 18.9 [15]	-918.6 ± 20.7 [2]	-948.2 ± 10.1	0.205	0.109
Pi10_all	4.0 ± 0.2 [79]	4.0 ± 0.1 [15]	4.0 ± 0.2 [2]	3.9 ± 0.1	0.297	0.713
Pi10_leq20	3.9 ± 0.1 [79]	3.9 ± 0.1 [15]	3.9 ± 0.1 [2]	3.9 ± 0.0	0.643	0.543
MRC scales (Level 1-5)	1.50 ± 0.68 [40]	1.46 ± 0.70 [8]	1.60 ± 0.90 [4]	2.00 ± 0.00	0.435	0.638
MRC scales ≥ 3, n (%)	47 (6.6) [40]	6 (7.1) [8]	6 (15.0) [4]	0 (0.0)	0.256	0.359
CAT score	7.4 ± 6.4 [15]	8.8 ± 7.4 [2]	7.7 ± 8.8	6.5 ± 2.1	0.34	0.226
SGRQ-Total score	14.1 ± 14.7 [37]	17.3 ± 18.4 [3]	16.8 ± 19.6 [5]	17.5 ± 0.8	0.531	0.178
Never-smokers	**n=401**	**n=55**	**n=15**	**n=2**		
Age (years; mean ± SD)	67.1 ± 9.8	64.6 ± 10.1	69.3 ± 10.1	75.0 ± 11.3	0.122	-
AAT levels (g/L)	1.28 ± 0.16 [2][a]	1.07 ± 0.14 [1][b]	0.88 ± 0.15 [1][c]	0.60 ± 0.08[c]	**<0.001**	**<0.001**
Sex – no. of patients (%)					0.843	
Male	206 (51.4)	26 (47.3)	9 (60.0)	1 (50.0)		
Female	195 (48.6)	29 (52.7)	6 (40.0)	1 (50.0)		
BMI (kg/m^2)	26.9 ± 4.8	27.2 ± 4.9	26.8 ± 3.1	23.0 ± 0.7	0.523	0.868

Post-bronchodilator FEV1, L	2.6 ± 0.8	2.7 ± 1.0	2.9 ± 0.9	2.6 ± 0.6	0.752	0.310
Post-bronchodilator FEV1 % predicted	96.9 ± 18.2	94.5 ± 18.7	104.0 ± 15.0	107.8 ± 30.5	0.262	0.467
Post-FEV1/FVC	72.3 ± 8.9	72.4 ± 8.4	74.2 ± 7.8	67.1 ± 14.8	0.901	0.698
FEV1 reversibility, %	4.7 ± 7.4	6.0 ± 7.9	3.2 ± 4.9	1.5 ± 2.1	0.282	0.526
FEF25-75	2.0 ± 1.0	2.0 ± 1.0	2.2 ± 0.9	1.6 ± 0.5	0.622	0.603
DLCO, % predicted	111.0 ± 22.9 [14][a]	119.6 ± 29.5 [4][a]	108.2 ± 22.1 [2]	64.9 ± 29.8[b]	**0.003**	**0.003**
RV, % predicted	113.9 ± 31.2 [19]	118.0 ± 30.1 [3]	102.1 ± 15.7	127.5 ± 13.4	0.208	0.461
FRC, % predicted	108.0 ± 24.8 [18]	112.0 ± 25.0 [3]	101.9 ± 14.8	128.5 ± 1.8	0.233	0.486
TLC, % predicted	116.9 ± 17.5 [17]	120.3 ± 22.2 [3]	114.3 ± 10.5	130.1 ± 2.8	0.409	0.459
LAA-950 (%)	3.6 ± 3.5 [56][a]	3.7 ± 4.3 [10][a]	3.1 ± 1.7[a]	25.1 ± 17.9[b]	**<0.001**	**<0.001**
PD15	-915.6 ± 20.9 [56][a]	-915.2 ± 21.4 [10][a]	-919.0 ± 10.9	-960.9 ± 20.3[b]	**0.020**	**0.032**
Pi10_all	3.9 ± 0.1 [56]	3.9 ± 0.1 [10]	4.0 ± 0.2	4.1 ± 0.2	0.419	0.272
Pi10_leq20	3.9 ± 0.1 [56]	3.8 ± 0.1 [10]	3.9 ± 0.1	4.0 ± 0.1	0.379	0.519
MRC scales (Level 1-5)	1.3 ± 0.6 [11]	1.3 ± 0.6 [3]	1.2 ± 0.4	1.5 ± 0.7	0.619	0.745
MRC scales ≥ 3, n (%)	11 (2.8) [11]	2 (3.8) [3]	0 (0.0)	0 (0.0)	0.791	0.973
CAT score	5.5 ± 4.7 [10]	6.4 ± 5.0 [1]	2.8 ± 3.3	5.0 ± 4.2	0.072	0.08
SGRQ-Total score	10.2 ± 12.5 [218]	10.4 ± 10.9 [23]	2.3 ± 1.5 [10]	25.0 ± . [1]	0.155	0.251

See footnotes of Table 3.

Table S3. Clinical characteristics by *SERPINA1* genotyping groups with MZ heterozygotes-only in group 3.

Characteristics	Group 1: no deficiency n=1,149	Group 2: mild deficiency n=147	Group 3: MZ n=40	Group 4: severe deficiency n=4	P value*	P value adj. age, sex, and study sites
Age (years; mean ± SD)	66.8 ± 9.7	64.9 ± 9.6	69.9 ± 8.0	68.3 ± 10.6	**0.028**	-
AAT levels (g/L)	1.31 ± 0.17 [12][a]	1.10 ± 0.14 [1][b]	0.88 ± 0.13 [2][c]	0.50 ± 0.21[d]	**<0.001**	**<0.001**
Sex – no. of patients (%)					0.307	-
Male	647 (56.3)	85 (57.8)	27 (67.5)	1 (25.0)		
Female	502 (43.7)	62 (42.2)	13 (32.5)	3 (75.0)		
BMI (kg/m^2)	27.6 ± 5.0 [1]	27.9 ± 5.3	27.2 ± 4.1	25.7 ± 4.8	0.735	0.917
Smoking status – no. of patients (%)					0.523	0.725
Never	401 (34.9)	55 (37.4)	9 (22.5)	2 (50.0)		
Former	561 (48.8)	70 (47.6)	26 (65.0)	2 (50.0)		
Current	187 (16.3)	22 (15.0)	5 (12.5)	0 (0.0)		
Pack-years (ever-smokers)	16.7 ± 22.5 [19]	16.9 ± 21.9 [1]	16.3 ± 19.7	11.7 ± 22.2	0.956	0.983
Post-bronchodilator FEV1, L	2.55 ± 0.81	2.62 ± 0.88	2.6 ± 0.9	2.18 ± 0.62	0.654	0.976
Post-bronchodilator FEV1 % predicted	91.5 ± 20.4	90.8 ± 19.7	89.5 ± 21.5	90.1 ± 27.8	0.948	0.852
Post-FEV1/FVC	69.5 ± 10.4	69.4 ± 9.8	66.3 ± 12.3	59.7 ± 13.0	0.166	0.107
FEV1 reversibility, %	5.4 ± 7.9	6.7 ± 8.9	6.6 ± 12.6	7.9 ± 12.9	0.496	0.282
FEF25-75	1.77 ± 1.04	1.79 ± 1.04	1.6 ± 0.8	1.05 ± 0.68	0.421	0.482
DLCO, % predicted	106.4 ± 24.3 [56][a]	109.9 ± 26.5 [11][a]	95.5 ± 21.2 [4][b]	69.4 ± 19.1[b]	**<0.001**	**<0.001**
RV, % predicted	120.1 ± 34.8 [61]	125.1 ± 34.9 [10]	112.9 ± 34.5 [2]	145.9 ± 33.1	**0.036**	0.089
FRC, % predicted	110.3 ± 25.6 [54]	113.6 ± 25.2 [10][a]	104.4 ± 26.0 [2][b]	136.9 ± 19.7	**0.014**	**0.032**
TLC, % predicted	118.1 ± 17.4 [52]	120.2 ± 18.3 [11]	116.1 ± 15.2 [2]	133.8 ± 11.9	0.116	0.117
LAA-950 (%)	4.30 ± 4.52 [135][a]	4.11 ± 4.23 [25][a]	5.00 ± 5.68 [1][a]	19.89 ± 12.18[b]	**0.008**	**<0.001**
PD15	-917.2 ± 21.5 [135][a]	-916.5 ± 19.8 [25][a]	-922.3 ± 18.2 [1][a]	-954.5 ± 15.0[b]	**0.007**	**0.002**
Pi10_all	3.96 ± 0.16 [135]	3.97 ± 0.14 [25]	4.01 ± 0.16	4.00 ± 0.14	0.16	0.204
Pi10_leq20	3.87 ± 0.11 [135]	3.88 ± 0.11 [25]	3.91 ± 0.12 [1]	3.90 ± 0.08	0.174	0.215
MRC scales (Level 1-5)	1.44 ± 0.65 [51]	1.40 ± 0.67 [11]	1.5 ± 0.9 [2]	1.75 ± 0.50	0.333	0.77
MRC scales ≥ 3, n (%)	58 (5.3) [51]	8 (5.9) [11]	5 (13.2) [2]	0 (0.0)	0.199	0.284
CAT score	6.8 ± 5.9 [25]	7.9 ± 6.7 [3]	6.3 ± 8.4	5.8 ± 2.9	**0.039**	0.122
SGRQ-Total score	13.3 ± 14.4 [255]	15.5 ± 17.0 [26]	14.6 ± 18.5 [10]	20.0 ± 4.4 [1]	0.336	0.324
Self-reported comorbidities, n (%)						
Hypertension	402 (35.0)	51 (34.7)	14 (35.0)	2 (50.0)	0.929	0.778
CVD (any CVD excluding Hypertension)	325 (28.3)	40 (27.2)	15 (37.5)	1 (25.0)	0.581	0.893
Diabetes	123 (10.7)	13 (8.8)	5 (12.5)	0 (0.0)	0.808	0.96
Asthma	267 (23.2)	32 (21.8)	13 (32.5)	1 (25.0)	0.478	0.29

Respiratory medications, n (%)						
SAMA/SABA	35 (3.0)	6 (4.1)	1 (2.5)	0 (0.0)	0.778	0.931
LABA± SAMA/SABA	3 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	1	1
LAMA± SAMA/SABA	10 (0.9)	3 (2.0)	1 (2.5)	0 (0.0)	0.19	0.478
LAMA+LABA± SAMA/SABA	71 (6.2)	11 (7.5)	2 (5.0)	0 (0.0)	0.844	0.962
ICS± SAMA/SABA	2 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1	1
LABA+ICS± SAMA/SABA	99 (8.6)	10 (6.8)	1 (2.5)	1 (25.0)	0.233	0.383
LAMA+ICS± SAMA/SABA	3 (0.3)	1 (0.7)	0 (0.0)	0 (0.0)	0.46	0.819
LAMA+LABA+ICS± SAMA/SABA	32 (2.8)	5 (3.4)	3 (7.5)	0 (0.0)	0.264	0.544
Any above medications	255 (22.2)	36 (24.5)	8 (20.0)	1 (25.0)	0.87	0.944

See footnotes of Table 3.

Table S4. Primers to sequence the *SERPINA1* gene

	Forward	**Reverse**	**Amplicon Size (bp)**
Exon 2	5'- AATGCATTGCCAAGGAGAGTT C -3'	**5'- CAAGTACTTGGCACAGGC** TG **-3'**	866
Exon 3	5'- CCTCAGTCCCAACATGGCTAA G -3'	**5'- AGGGATGTGTGTCGTCAA** GG **-3'**	460
Exon 4	5'- CTTCCTCAGCCTCAGGAC AG -3'	**5'- AGTTCCCATCTTAGTGTGGGT** G **-3'**	376
Exon 5	**5'- GGATTTACAGTCACATGCAGG** C **-3'**	5'- GGGAGTGAGCGCTTC CTG -3'	387

Sequencing primers are shown in bold.

Table S5. Primers for allele-specific PCR

AS-PCR specificity	Allele-Specific Forward Primer sequence	Common Reverse Primer sequence	Size of the amplicon	PCR conditions*
rs1303 Glu376Asp	5'- CATGAAGAGGGGAGACTTGGTATTTTG**G** -3'	5'- CTGAGTTCGCCTTCAGCCTATACC-3'	4598 bp	**A**
	5'- CATGAAGAGGGGAGACTTGGTATTTTG**T** -3			
rs1303 Glu376Asp	5'- CATGAAGAGGGGAGACTTGGTATTTTG**G** -3'	5'- CAAGTACTTGGCACAGGCTGG -3'	4888 bp	**A**
	5'- CATGAAGAGGGGAGACTTGGTATTTTG**T** -3'			
rs1303 Glu376Asp	5'- CATGAAGAGGGGAGACTTGGTATTTTG**G** -3'	5'- AGGGATGTGTGTCGTCAAGG -3'	2789 bp	**A**
	5'- CATGAAGAGGGGAGACTTGGTATTTTG**T** -3'			
rs1379209512 Leu-2Gln	5'- CCTGGTCCCTGTCTCCC**A** -3'	5'- CCTCAGTCCCAACATGGCTAAG -3'	2382 bp	**B**
	5'- CCTGGTCCCTGTCTCCC**T** -3'			
rs759736224 Pro255Thr	5'- GTGCTGTAGTTTCCCCTCATCAG**T** -3'	5'- CTGGCTGAGTTCGCCTTCAG -3'	2149 bp	**C**
	5'- GTGCTGTAGTTTCCCCTCATCAGG -3'			
rs766025736 Arg101Cys	5'- TGAAGGCTTCCAGGAACTCCTC**T** -3	5'- CCTCAGTCCCAACATGGCTAAG -3	2082 bp	**D**
	5'- TGAAGGCTTCCAGGAACTCCTC**C** -3			
rs28929474 Glu342Lys	5'- CCAGCAGCTTCAGTCCCTTTCT**T** -3'	5'- AGGGATGTGTGTCGTCAAGG -3'	2680 bp	**E**
	5'- CCAGCAGCTTCAGTCCCTTTCT**C** -3'			
rs745463238 Ser14Phe	5'- GCTGCCCAGAAGACAGATACAT**T** -3'	5'- CCTCAGTCCCAACATGGCTAAG -3'	2342 bp	**F**
	5'- GCTGCCCAGAAGACAGATACAT**C** -3'			
rs9630 Ala332Ala	5'- GTCAGCACAGCCTTATGCAC**A** -3'	5'- AGGGATGTGTGTCGTCAAGG -3'	2650 bp	**D**
	5'- GTCAGCACAGCCTTATGCAC**G** -3'			
rs1049800 Asp256Asp	5'- AGGTGCTGTAGTTTCCCCTC**G** -3'	5'- CTGAGTTCGCCTTCAGCCTATACC -3'	2148 bp	**A**
	5'- CAGGTGCTGTAGTTTCCCCTC**A** -3'			
rs200414579 Gly167Gly	5'- GCTCCTTGACCAAATCCACAATTTT**T** -3'	5'- CTGAGTTCGCCTTCAGCCTATACC -3'	435 bp	**G**
	5'- GCTCCTTGACCAAATCCACAATTTT**C** -3'			
rs147283849 Leu-10Leu	5'- ATCCTCCTGCTGGCAGGC**T** -3'	5'- CCTCAGTCCCAACATGGCTAAG -3'	2408 bp	**A**
	5'- ATCCTCCTGCTGGCAGGC**C** -3'			

*see Table S6.

Table S6. AS-PCR cycling conditions

PCR conditions	Taq polymerase	Initial Denaturation	Denaturation	Annealing Temperature	Elongation
A	NEB Q5	98°C, 30 sec	98°C, 10 sec	68°C, 30 sec	72°C, 3 min
B	NEB Q5	98°C, 30 sec	98°C, 10 sec	69°C, 30 sec	72°C, 3 min
C	NEB Q5	98°C, 30 sec	98°C, 10 sec	70°C, 30 sec	72°C, 3 min
D	NEB Q5	98°C, 30 sec	98°C, 10 sec	72°C, 30 sec	72°C, 3 min
E	HotStarTaq	95°C, 15 min	94°C, 15 sec	61.5°C, 40 sec	72°C, 2.5 min
F	HotStarTaq	95°C, 15 min	94°C, 15 sec	63°C, 40 sec	72°C, 2.5 min
G	HotStarTaq	95°C, 15 min	94°C, 15 sec	57°C, 30 sec	72°C, 1 min

A, B, C and **D**: Final reaction volume of 50 µL, 1X PCR buffer, 1X Q5 GC enhancer buffer, 200 µM of each dNTP, 500 nM of each primer, 2.5 Units of Taq Polymerase and 35 PCR cycles of denaturation-annealing-elongation.

E and **F**: Final reaction volume of 50 µL, 1X PCR buffer, 1X Q-Solution, 160 µM of each dNTP, 400 nM of each primer, 2 Units of Taq Polymerase and 35 PCR cycles of denaturation-annealing-elongation.

G: Final reaction volume of 25 µL, 1X PCR buffer, 1X Q-Solution, 160 µM of each dNTP, 200 nM of each primer, 1 Units of Taq Polymerase and 35 PCR cycles of denaturation-annealing-elongation.

NEB Q5: Q5® High-Fidelity DNA Polymerase, New England Biolabs Ltd., Ontario, Canada

HotStarTaq: HotStarTaq DNA polymerase, Qiagen, Ontario, Canada

Table S7. Frequencies, allelic background, AAT serum levels, and lung phenotypes among carriers of rare genetic variants with no Pi typing.

Exon	rs #	Protein	Background	# of carriers	AAT levels	DLCO, % predicted	LAA-950
2	rs1343069141	Ser-19Ser	M3	1	1.22	NA	4.44
2	rs147283849	Leu-10Leu	M1 (Ala^{213})	3	1.06±0.03	133.24±33.09	3.59±1.48 [1]
2	rs1379209512	Leu-2Gln	M1 (Ala^{213})	2	1.27±0.06	91.34±35.24	2.32 [1]
2	rs150784949	Phe33Phe	M1	2	1.17±0.10	103.72±14.21	4.19 [1]
2	rs113817720	Ala58Thr	M1	1	1.12	129.44	2.39
2	rs758820515	Pro88Leu	M1 (Ala^{213})	1	1.00	136.63	2.43
2	rs766025736	Arg101Cys	S	1[a]	0.99	146.27	4.20
2	rs1344951022	Thr113Thr	M1	1	1.24	NA	2.75
2	rs200414579	Gly167Gly	M3	1	1.39	112.26	7.81
3	New1	Lys222Lys	M1	1	1.32	96.62	1.30
3	rs759736224	Pro255Thr	M3	1	1.38	72.48	1.39
3	rs1049800	Asp256Asp	M1 or M1 (Ala^{213})[b]	6	1.27±0.17	110.16±17.70	8.52±10.72
3	New2	Lys274Asn	M1 (Ala^{213})	1	1.24	101.40	0.55
4	rs141620200	Ala284Ser	M1 (Ala^{213})	11	1.20±0.21 [1]	96.11±17.97 [2]	3.32±4.26 [1]
4	rs139964603	Val302Ile	M1	1	1.16	104.97	12.98
4	New3	Leu318Phe	M1	1	1.50	97.17	3.29
5	rs9630	Ala332Ala	M1 or M1 (Ala^{213})[c]	2	1.25±0.16	117.53 [1]	2.84±1.17
5	rs201318727	Ile340Val	R	1	1.34	114.92	2.82

The number of missing values is indicated in square brackets

[a]MS

[b]Background M1 for individuals of Asian ancestry (n=5) and M1 (Ala^{213}) for European ancestry (n=1).

[c]Background M1 for one individual of European ancestry and M1 (Ala^{213}) for African ancestry (n=1).

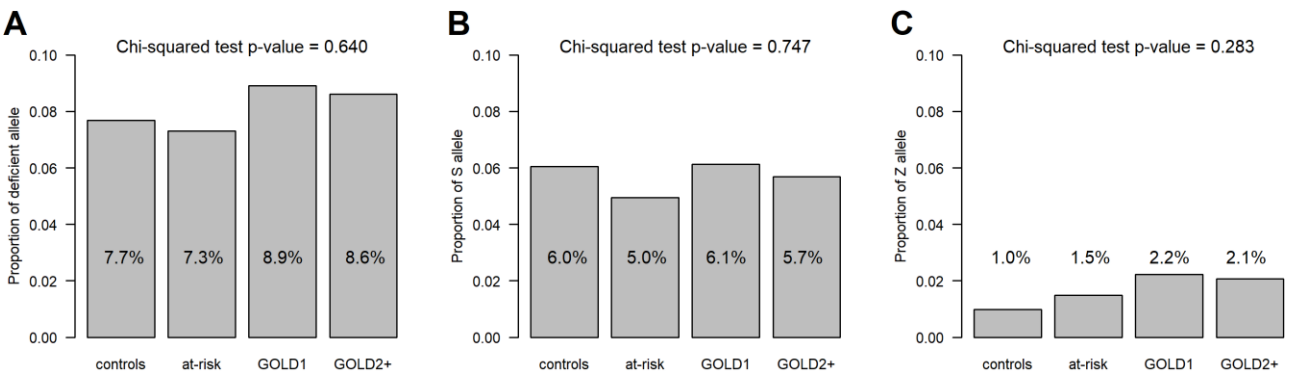


Figure S1. Proportion of all deficient alleles (A), S allele (B), and Z allele in CanCOLD subgroups including healthy control free of smoking history and airway obstruction (controls), ever-smokers free of airway obstruction (at-risk), individuals with mild COPD (GOLD1) and individuals with moderate-severe COPD (GOLD2+).

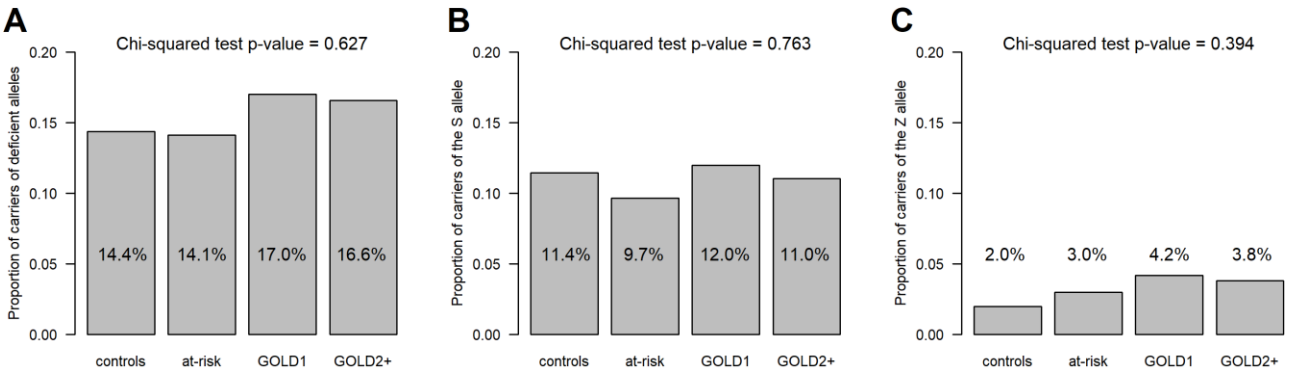


Figure S2. Proportion of carriers of deficient alleles (A), S allele (B), and Z allele in CanCOLD subgroups including healthy control free of smoking history and airway obstruction (controls), ever-smokers free of airway obstruction (at-risk), individuals with mild COPD (GOLD1) and individuals with moderate-severe COPD (GOLD2+).