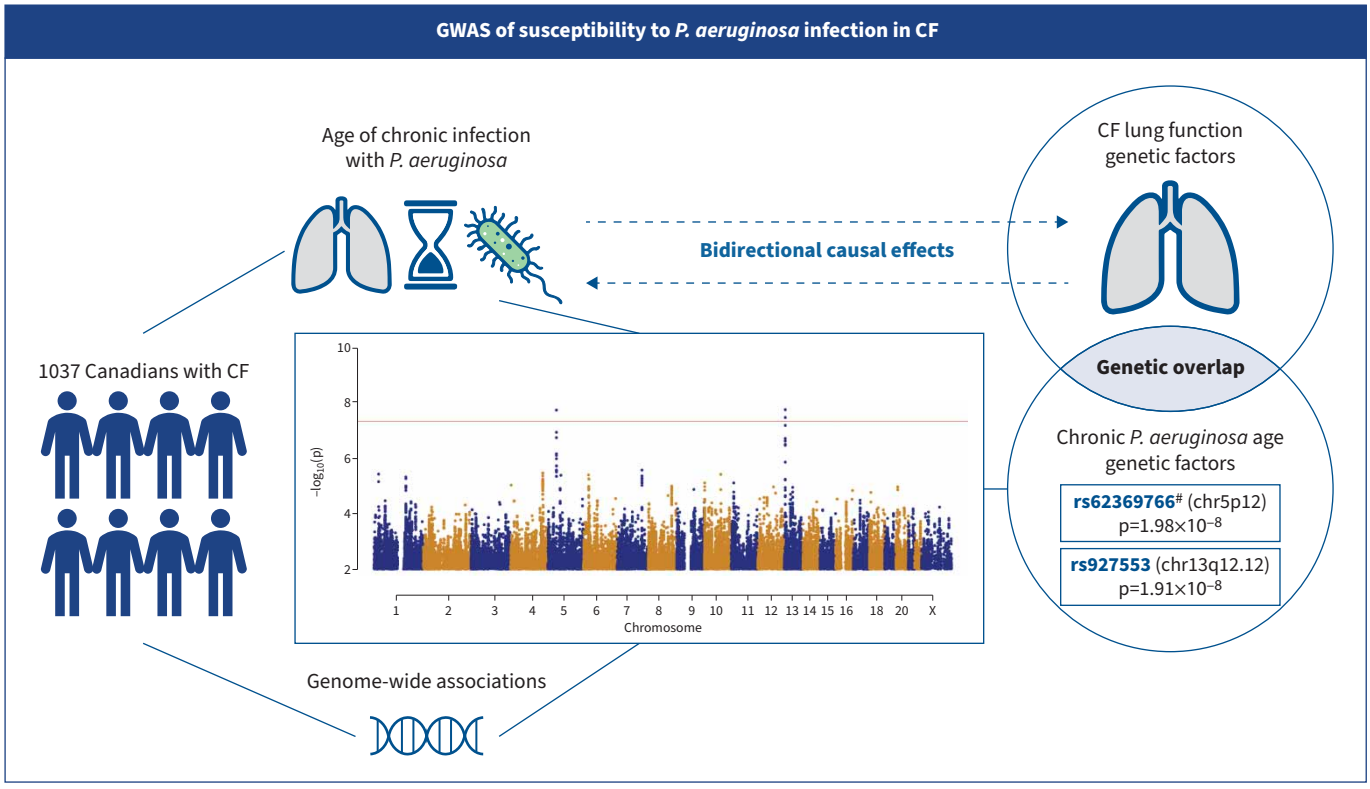
Genome-wide association study of susceptibility to *Pseudomonas aeruginosa* infection in cystic fibrosis

Boxi Lin, Jiafen Gong, Katherine Keenan, Fan Lin, Yu-chung Lin, Julie Mésinèle, Claire Calmel, Badreddine Mohand Oumoussa, Pierre-Yves Boëlle, Loïc Guillot, Harriet Corvol, Valerie Waters, Lei Sun and Lisa J. Strug



GRAPHICAL ABSTRACT Summary of the study. In a genome-wide association study (GWAS) of 1037 Canadians with cystic fibrosis (CF), we found two novel loci linked to chronic *Pseudomonas aeruginosa* infection age. We found evidence of a shared polygenic component and a potential causal relationship between chronic *P. aeruginosa* infection and lung disease. #: the rs62369766 locus was validated using an independent French cohort (n=501).

https://doi.org/10.1183/13993003.00062-2024

Genome-wide association study of susceptibility to *Pseudomonas aeruginosa* infection in cystic fibrosis

Boxi Lin[1,2], Jiafen Gong[2], Katherine Keenan[2], Fan Lin[2], Yu-chung Lin[1,2], Julie Mésinèle[3,4], Claire Calmel[3], Badreddine Mohand Oumoussa[5], Pierre-Yves Boëlle[6], Loïc Guillot[3], Harriet Corvol[3,7], Valerie Waters[8,9], Lei Sun[1,10] and Lisa J. Strug[1,2,10,11,12]

[1]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [2]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada. [3]Sorbonne Université, Inserm U938, Centre de Recherche Saint-Antoine (CRSA), Paris, France. [4]Inovarion, Paris, France. [5]Sorbonne Université, Inserm, UMS Production et Analyse des données en Sciences de la vie et en Santé (PASS), Plateforme Post-génomique de la Pitié-Salpêtrière, Paris, France. [6]Sorbonne Université, Inserm, Institut Pierre Louis d'Epidémiologie et de Santé Publique (iPLESP), AP-HP, Hôpital Saint-Antoine, Paris, France. [7]Sorbonne Université, AP-HP, Hôpital Trousseau, Service de Pneumologie Pédiatrique, Paris, France. [8]Division of Infectious Diseases, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada. [9]Translational Medicine Research Program, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada. [10]Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada. [11]Department of Computer Science, University of Toronto, Toronto, ON, Canada. [12]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada.

Corresponding author: Lisa J. Strug (lisa.strug@utoronto.ca)

Shareable abstract (@ERSpublications)
This GWAS on 1037 Canadians with CF found two novel loci linked to chronic *Pseudomonas aeruginosa* (*Pa*) infection age, along with evidence of a shared polygenic component and a potential causal relationship between chronic *Pa* infection and lung disease https://bit.ly/4fcMeFg

Cite this article as: Lin B, Gong J, Keenan K, *et al*. Genome-wide association study of susceptibility to *Pseudomonas aeruginosa* infection in cystic fibrosis. *Eur Respir J* 2024; 64: 2400062 [DOI: 10.1183/13993003.00062-2024].

This article has an editorial commentary: https://doi.org/10.1183/13993003.01224-2024

Received: 10 Oct 2022
Accepted: 10 July 2024

Abstract

Background *Pseudomonas aeruginosa* is a common pathogen that contributes to progressive lung disease in cystic fibrosis (CF). Genetic factors other than CF-causing *CFTR* (CF transmembrane conductance regulator) variations contribute ∼85% of the variation in chronic *P. aeruginosa* infection age in CF according to twin studies, but the susceptibility loci remain unknown. Our objective is to advance understanding of the genetic basis of host susceptibility to *P. aeruginosa* infection.

Materials and methods We conducted a genome-wide association study of chronic *P. aeruginosa* infection age in 1037 Canadians with CF. We subsequently assessed the genetic correlation between chronic *P. aeruginosa* infection age and lung function through polygenic risk score (PRS) analysis and inferred their causal relationship through bidirectional Mendelian randomisation analysis.

Results Two novel genome-wide significant loci with lead single nucleotide polymorphisms (SNPs) rs62369766 (chr5p12; $p=1.98\times10^{-8}$) and rs927553 (chr13q12.12; $p=1.91\times10^{-8}$) were associated with chronic *P. aeruginosa* infection age. The rs62369766 locus was validated using an independent French cohort (n=501). Furthermore, the PRS constructed from CF lung function-associated SNPs was significantly associated with chronic *P. aeruginosa* infection age (p=0.002). Finally, our analysis presented evidence for a causal effect of lung function on chronic *P. aeruginosa* infection age (β=0.782 years, $p=4.24\times10^{-4}$). In the reverse direction, we observed a moderate effect (β=0.002, p=0.012).

Conclusions We identified two novel loci that are associated with chronic *P. aeruginosa* infection age in individuals with CF. Additionally, we provided evidence of common genetic contributors and a potential causal relationship between *P. aeruginosa* infection susceptibility and lung function in CF. Therapeutics targeting these genetic factors may delay the onset of chronic infections, which account for significant remaining morbidity in CF.

Introduction

Pseudomonas aeruginosa is a common respiratory pathogen in cystic fibrosis (CF) [1–3]. Chronic respiratory infection by *P. aeruginosa* is associated with more rapid lung function decline [4–6] and reduced survival in

CF [7, 8]. Up to 60% of individuals with CF are eventually infected with *P. aeruginosa* [1, 9, 10], but the ages of onset of the first and of chronic *P. aeruginosa* infection, defined as three consecutive years of culture data with at least one positive culture in at least two of the three years, vary considerably among individuals who have the same causal *CFTR* (CF transmembrane conductance regulator) variants, indicating the presence of modifying factors [11]. Host genetic factors have been shown to play a significant role in the establishment and timing of *P. aeruginosa* infection in CF [11, 12]. The *CFTR* functional class has been identified as an important risk factor for the age of *P. aeruginosa* acquisition [13]. Moreover, the estimated heritability of age of onset of chronic *P. aeruginosa* infection is ∼85% from twin studies [11], but the contributing genetic modifiers and pathways remain unknown. In contrast, there is less evidence from these same twin studies that the age at first infection with *P. aeruginosa* is heritable, guiding genetic studies to focus on the analysis of chronic *P. aeruginosa* infection age [11].

Discovery of genetic modifiers affecting age of onset of *P. aeruginosa* infection could inform contributing mechanisms to host *P. aeruginosa* infection susceptibility in CF. While eradication therapy has already been employed to target early *P. aeruginosa*, genetic modifiers may further serve as biomarkers for screening CF subgroups that could benefit from earlier intervention with aggressive infection-targeting treatment. Although an exome sequencing study [14] and candidate gene studies [15] have identified potential genetic modifiers for phenotypes related to age of *P. aeruginosa* infection [12], to the best of our knowledge there are no genetic association studies for age of chronic *P. aeruginosa* infection genome-wide and there is a lack of independent assessment of previously reported loci from genetic studies of *P. aeruginosa* infection.

Additionally, we are interested in understanding the genetic overlap and causal relationship between age of *P. aeruginosa* infection and lung disease severity in CF. A previous study has shown that earlier *P. aeruginosa* infection, particularly before 5 years of age, is strongly associated with severe CF lung disease later in life [16]. Such evidence led us to consider the effect of earlier *P. aeruginosa* infection on pulmonary function in individuals with CF. In particular, we are interested in investigating whether the phenotypic correlation between chronic *P. aeruginosa* infection age and lung disease severity is driven by a common genetic architecture, which could guide future research in choosing therapeutic targets to delay the onset of chronic *P. aeruginosa* infections that will have the greatest impact on improving lung disease.

In participants of the Canadian CF Gene Modifier Study (CGMS), we performed a genome-wide association study (GWAS) and identified two novel loci associated with age of chronic *P. aeruginosa* infection in CF. We validated these findings in an independent cohort from France. We further demonstrated genetic overlap and a potential causal relationship between age of chronic *P. aeruginosa* infection and lung function in CF. Some of the results in this paper have been previously reported in the form of an abstract [17].

Material and methods

Please see the supplementary material for more details.

Recruitment and study design

The study included a total of 2728 genotyped individuals with CF from CGMS and an independent validation cohort from the French CF Modifier Gene Study (FCGS) (n=501). To minimise the confounding effect of variation in severity of the causal *CFTR* genotype, we included only individuals with genotypes known to confer pancreatic insufficiency [2]. Moreover, to reduce the confounding influence of CF treatment, we restricted the analysis to modulator-naïve individuals. We obtained written informed consent from adults, or from parents or guardians for participants aged <18 years old [2]. CGMS was approved by the Hospital for Sick Children Research Ethics Board with research ethics approval number 1000065760. FCGS was approved by the French ethical committees and written informed consent was obtained from each patient and/or guardian.

The same sample quality control procedures were implemented for the CGMS and FCGS analyses (supplementary appendix S1) and the phenotype analysed in the FCGS conformed to the inclusion criteria specified in a previous study using the same cohort [6]. The flowcharts of subject inclusion and exclusion are outlined in figure 1.

Genotype data

Genotyping, variant calling and quality control were conducted as in Gong *et al.* [2], with phasing and imputation as described in Panjwani *et al.* [18]. After standard quality control and imputation, a total of 6 994 213 genotyped and imputed SNPs (with imputation quality allelic $r^2 \geqslant 0.8$) were analysed. We then

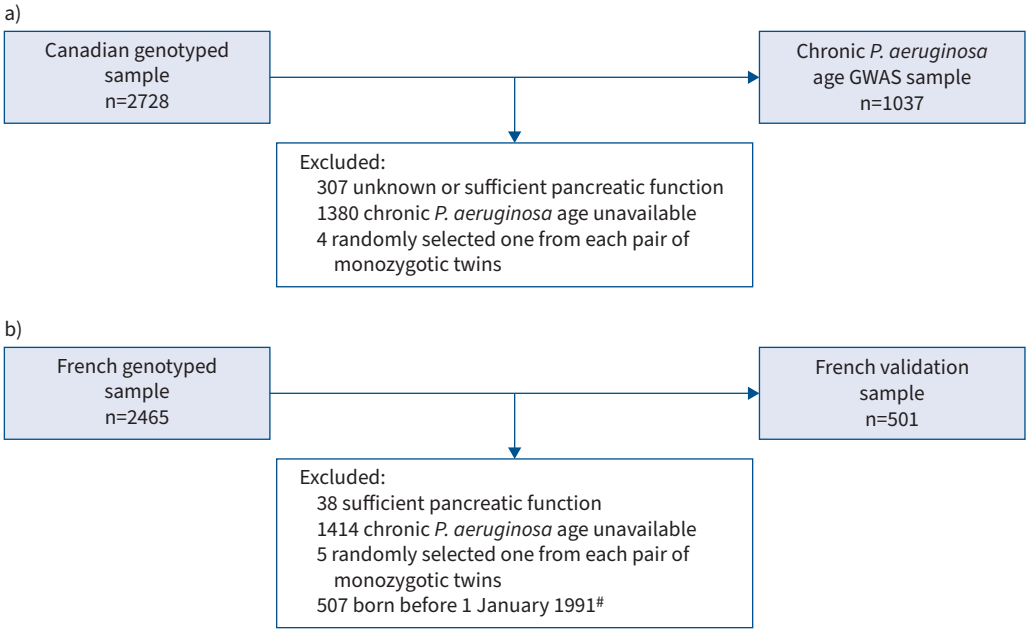


FIGURE 1 Flowcharts illustrating inclusion and exclusion criteria for a) 2728 genotyped individuals with cystic fibrosis (CF) from the Canadian CF Gene Modifier Study for the genome-wide association study (GWAS) of chronic *Pseudomonas aeruginosa* age and b) 2465 genotyped individuals from the French CF Modifier Gene Study for validation analysis. To minimise the confounding effect of variation in severity of the causal *CFTR* (CF transmembrane conductance regulator) genotype, we included only individuals with genotypes known to confer pancreatic insufficiency [2]. Moreover, to reduce the confounding influence of CF treatment, we restricted the analysis to modulator-naïve individuals. #: we followed the sample inclusion procedures in Mesinele *et al.* [6] which restricted to individuals born between 1 January 2001 and 31 December 2019. However, within this cohort, only 144 individuals (born before 2010) were genotyped and presented chronic *P. aeruginosa* infection. To ensure a sufficient sample size, we instead utilised a cohort (n=501) born within a two-decade span preceding the date of birth of the youngest genotyped individual. This selected cohort consists of individuals born between 1 January 1991 and 31 December 2009.

conducted principal component analysis to control for population stratification and inferred sample relatedness. Details are provided in supplementary appendix S1 and supplementary figures Si–Siii.

Phenotype definitions

The *P. aeruginosa* culture data in the Canadian study were obtained from chart review and registry data. Specifically, we defined the phenotype based on longitudinal data that were collected as part of the national registry, with clinics conducting chart reviews to address data inconsistencies where possible. Following Green *et al.* [11], the first infection was defined as the first positive respiratory culture for *P. aeruginosa*, and chronic infection was defined as the presence of three consecutive years of records with at least one positive culture in at least two of the three years. The age of the subject at the time of the first culture meeting this criterion was recorded as the age of chronic infection (supplementary appendix S1 and supplementary figure Siv; see Discussion) [11]. In the French study, chronic infection was defined as at least three positive samples at least 1 month apart over a 6-month period, and chronic *P. aeruginosa* age was the age at the third isolation [6]. Calculating the phenotype defined by Green *et al.* [11] using the French data is not possible due to the lack of longitudinal microbiological analysis of sputum samples.

The distributions of the first *P. aeruginosa* age and chronic *P. aeruginosa* age were highly right-skewed (supplementary appendix S1 and supplementary figure Sv), which could reduce the power of the association test [19]. The residuals, after regressing on current age (supplementary appendix S7), appeared symmetric (Kolmogorov–Smirnov symmetric test p=0.223; supplementary appendix S1 and supplementary figure Sv) [20]. Thus, these residuals are the primary outcomes for our genetic analyses (supplementary appendix S1).

Heritability estimation

We used GCTA [21] to estimate the SNP-based heritability (h^2) of the two residualised phenotypes (supplementary appendix S1). The estimation was conducted on unrelated participants of European origin with CF from CGMS (n=937 for chronic *P. aeruginosa* and n=1476 for first *P. aeruginosa* age).

Association testing

Genome-wide association testing was carried out using a linear mixed effects model implemented in the GENESIS package [22] to account for sibling relationships in CGMS (supplementary appendix S1). For both the first and chronic *P. aeruginosa* infection age, we conducted genome-wide association analysis while controlling for sex and genetic principal components. We conducted a discovery GWAS on individuals of European ancestry. We then validated the GWAS with the remaining independent non-European participants. Next, we jointly analysed both European individuals and non-European individuals to maximise power. We used the conventional p-value threshold of 5×10^{-8} to declare genome-wide significance [23]. Finally, we validated the association studies at the identified genome-wide significant signals using the independent French cohort (n=501). We used a Bonferroni corrected threshold of 0.025 at each of the two genome-wide significant SNPs to conclude significance at the 5% level.

Cross-trait polygenic risk score

We conducted a cross-trait polygenic risk score (PRS) analysis to assess the genetic overlap between chronic *P. aeruginosa* age and the International CF Gene Modifier Consortium lung phenotype (SaKnorm, a measurement of lung disease based on forced expiratory volume in 1 s (FEV_1) adjusted for age, height, sex and CF-related mortality, was calculated prior to taking *CFTR* modulators [24, 25]). We first selected 7756 top-ranked and independent (linkage disequilibrium (LD) $r^2<0.1$) SNPs associated with SaKnorm (with $p<0.014$, determined from a grid search for the optimal threshold with the corresponding PRS explaining the largest proportion of chronic *P. aeruginosa* age variation; supplementary appendix S2), based on the summary statistics from Corvol *et al.* [24] (n=6365). We then constructed the PRS using the 120 CGMS individuals who were not included in the consortium lung function GWAS. Subsequently, we estimated the proportion of variance in chronic *P. aeruginosa* age that could be explained by the PRS of SaKnorm through regression (supplementary appendix S2).

Bidirectional Mendelian randomisation

We conducted bidirectional Mendelian randomisation (MR) analysis [26] to investigate the causal relationship between chronic *P. aeruginosa* infection and lung function in CF, utilising the summary statistics of our chronic *P. aeruginosa* age GWAS (n=1037) and the SaKnorm GWAS in Corvol *et al.* [24] (n=6365). We first tested whether the onset of chronic *P. aeruginosa* infection affects lung function in CF. To improve the statistical power, we used the set of 677 independent SNPs (LD $r^2<0.001$) as the multivariate instrument variables (IVs) for chronic *P. aeruginosa* age, the exposure variable. We chose this set based on a PRS that explained the largest proportion of chronic *P. aeruginosa* age variation (supplementary appendices S2 and S3). Reversely, following the same steps, we tested whether reduced lung function leads to earlier chronic *P. aeruginosa* age. Sensitivity analysis (supplementary appendix S3) was conducted with different multivariate IV selection criteria, as well as alternative MR methods accounting for bias arising from weak IVs (MR weighted median [27]), outliers (MR weighted median and simple mode) and horizontal pleiotropy (MR-Egger [28]).

Results

Study samples

The demographic and clinical characteristics of the study samples are detailed in table 1. The CGMS sample included 2728 genotyped individuals born between 1940 and 2020, with the majority (mean age±3SD) born between the 1970s and 1990s. 73.3% of individuals were diagnosed with CF before the age of 3 years. The primary enrolment periods were from 2002 to 2006 and 2012 to 2018 (supplementary figures S3 and S4). For the GWAS, we excluded individuals with missing response variables (details and proportions of reasons for missing data are available in supplementary appendix S6 and supplementary figure Si), resulting in 1653 individuals for the age of first *P. aeruginosa* infection analysis and 1037 individuals for the age of chronic *P. aeruginosa* infection analysis. The clinical attributes of the first and chronic *P. aeruginosa* age GWAS samples were largely consistent (table 1). In comparison, while the French validation cohort was younger than the Canadian cohort, the mean first and chronic *P. aeruginosa* ages and proportions of individuals diagnosed with CF before the age of 3 years (83.6%) remained consistent across the cohorts (table 1).

Heritability estimates

The estimated SNP-based heritability was 22.5% for the first *P. aeruginosa* infection age residual (p=0.090) compared to 46.0% for the chronic *P. aeruginosa* infection age residual (p=0.044)

TABLE 1 Demographic and clinical characteristics of the Canadian CF Gene Modifier Study (CGMS) and French CF Modifier Gene Study (FCGS) individuals

	CGMS genotyped[#]	First *P. aeruginosa* age GWAS[¶]	Chronic *P. aeruginosa* age GWAS[+]	Chronic *P. aeruginosa* age FCGS validation analysis[§]
Individuals (n)	2728	1653	1037	501
Age[f] (years)				
Mean±SD	34.6±13.6	34.0±11.9	36.3±11.2	23.6±4.7
Range	(3.5–90.1)	(5.0–80.5)	(6.8–71.1)	(12.6–30.7)
Male (%)	53.3	51.8	51.7	48.9
Age at CF diagnosis (years)	4.1±8.8	2.3±5.1	2.1±4.1	1.3±2.4
***CFTR* genotype distribution (%)**				
F508del/F508del	52.1	58.3	59.0	66.7
F508del/MF	37.8	33.6	33.1	31.3
MF/MF	10.2	8.2	7.9	2.0
Pancreatic exocrine insufficient[##] (%)	88.7	100	100	100
Sweat chloride ($ng \cdot mL^{-1}$)	97.4±22.5	101.2±19.6	102.0±20.3	102.1±26.7
Newborn screening (%)	11.6	8.8	11.4	22.2
European ancestry (%)	96.0	96.1	95.7	92.2
First *P. aeruginosa* age (years)				
Mean±SD	9.9±10.1	8.8±8.2	9.0±7.8	7.3±6.1
Range	(0.0–74.6)	(0.0–64.4)	(0.0–52.1)	(0.0–28.4)
Range of diagnosis date	06/11/1966–28/12/2018	06/11/1966–28/12/2018	06/11/1966–29/08/2017	12/07/1991–13/08/2021
Missing rate (%)	33.6	NA	7.9	0
Chronic *P. aeruginosa* age (years)				
Mean±SD	12.7±9.3	11.4±7.9	11.9±7.9	11.6±6.3
Range	(0.3–66.8)	(0.3–52.1)	(0.3–52.1)	(0.01–29.2)
Range of diagnosis date	06/11/1966–29/08/2017	06/11/1966–29/08/2017	06/11/1966–29/08/2017	01/09/1991–13/08/2021
Missing rate (%)	59.1	42.3	NA	NA

Data are presented as mean±SD, unless otherwise stated. CF: cystic fibrosis; *P. aeruginosa*: *Pseudomonas aeruginosa*; GWAS: genome-wide association study; *CFTR*: CF transmembrane conductance regulator; MF: minimal function; NA: not applicable. [#]: all individuals from CGMS who passed quality control; [¶]: sample from CGMS for the GWAS of the first *P. aeruginosa* age, where individuals have insufficient pancreatic function and non-missing first *P. aeruginosa* age and are modulator-free; [+]: sample from CGMS for the GWAS of the chronic *P. aeruginosa* age, where individuals have insufficient pancreatic function and non-missing chronic *P. aeruginosa* age and are CF modulator-free; [§]: French sample for the validation analysis of the two genome-wide significant single nucleotide polymorphisms identified from the chronic *P. aeruginosa* age GWAS, where individuals have non-missing chronic *P. aeruginosa* age and insufficient pancreatic function and were born after 1 January 1991; [f]: calculated by 1 November 2021 minus the date of the birth; [##]: for the CGMS genotyped samples, 88.7% of individuals in the cohort were pancreatic insufficient; however, for the GWAS presented here of the first and chronic *P. aeruginosa* age, we only included individuals who were pancreatic insufficient.

(supplementary appendix S1 and supplementary table Si). These results are qualitatively consistent with the heritability estimates obtained from a previous twin study: 0% for first *P. aeruginosa* age and 85% for chronic *P. aeruginosa* age [11].

Since the heritability estimate of the first *P. aeruginosa* age residual was relatively low and did not reach statistical significance at the 5% level, our primary analyses focused on the chronic *P. aeruginosa* age residual. For completeness, results for a GWAS of the first *P. aeruginosa* age residual are available in supplementary appendix S4 and summary statistics can be found at https://github.com/strug-hub/gwas_psa.

Two novel loci associated with the chronic P. aeruginosa *age residual*

The GWAS of the chronic *P. aeruginosa* age residual was first conducted on 991 individuals of European ancestry and 6 994 213 SNPs passing quality control (supplementary appendix S1). We identified one genome-wide significant locus associated with the chronic *P. aeruginosa* age residual (figures 2 and 3 and table 2), with the top SNP rs927553 (β=1.53 years, $p=3.33\times10^{-8}$ at chr13q12.12). Additionally, we identified a second locus with the top SNP rs62369766, just shy of the genome-wide significant threshold (β= −1.96 years, $p=7.13\times10^{-8}$ at chr5p12). The two top SNPs from the loci are common (table 2) and jointly explain 3% of the phenotypic variation.

We found supporting evidence for the two loci with the remaining 46 non-European individuals. Despite the limited sample size of the non-European subset, the effect sizes and directions of the top-ranked SNPs were

consistent with the estimates from the European sample (table 2). To maximise power, we jointly analysed European and non-European individuals in a combined sample of 1037 individuals which provided genome-wide significance of the top SNPs at the two loci ($p=1.98\times10^{-8}$ at rs62369766 in chr5p12 and $p=1.91\times10^{-8}$ at rs927553 in chr13q12.12) (table 2). The variant rs62369766 is associated with an earlier onset of chronic *P. aeruginosa* infection by 2.00 years when carrying an additional copy of the effect allele T. Conversely, the variant rs927553 demonstrates a delay in the onset of chronic *P. aeruginosa* infection by 1.51 years for each additional copy of the effect allele G carried. The association of rs927553 with chronic *P. aeruginosa* infection showed a consistent effect direction, but it did not display a significant association at the 5% level in the validation study with the French cohort (β=0.05 years, p=0.90). The association of rs62369766 was validated in the independent French cohort (β= −1.24 years, for effect allele T *versus* the reference allele G, p=0.01; n=501) with a consistent effect direction as estimated in the Canadian sample (table 2).

Sensitivity analysis showed that the signals were robust when analyses were stratified based on potential confounding factors including CF diagnostic age, *CFTR* genotype, birth cohorts, recruitment sites and diagnosis by newborn screening (supplementary appendix S5). The signals also remained significant when adjusting for survival bias. This adjustment included 534 individuals whose chronic infection status was censored at the end of follow-up and 834 individuals who were chronically infected with *P. aeruginosa* before enrolment but lacked a specific date defining chronic infection (supplementary appendix S6 and supplementary figure Si; diagnostic plots: supplementary appendix S6 and supplementary figures Sii–Siv). Notably, the effects of the two GWAS signals diminished in the analysis that restricted to more recent cohorts (supplementary appendix S5 and supplementary table Siii), but SNP contributions remained in a consistent direction, suggesting a potential attenuating genetic effect possibly with the advances of CF treatment and interventions.

Chronic P. aeruginosa *age-associated locus is associated with lung function in non-CF populations*

We conducted a phenome-wide association study (PheWAS) using the GWAS Atlas [29] (Release 3: v20191115; accessed on 2 May 2023). We assessed the leading SNPs from the chr5p12 and chr13q12.12 loci with a p-value threshold of 1.05×10^{-5} determined by Bonferroni correction for analysing the 4756 phenotypes at the nominal family-wise error rate of 0.05. rs62369766 (chr5p12) was significantly associated with lung function measurements including forced vital capacity (FVC) ($p=2.59\times10^{-6}$) and FEV_1 ($p=3.81\times10^{-6}$) (supplementary appendix S1 and supplementary figure Svi) in the earlier genome-wide meta-analysis (n=400 102) of lung function in non-CF individuals of European descent from the UK Biobank (n=321 047) and SpiroMeta consortium [30] (n=79 055). Interestingly, both datasets are population-based and not CF-specific. For the four lung function phenotypes investigated (FEV_1, FVC, peak expiratory flow and FEV_1/FVC), rs62369766 demonstrated genetic effect directions consistent with

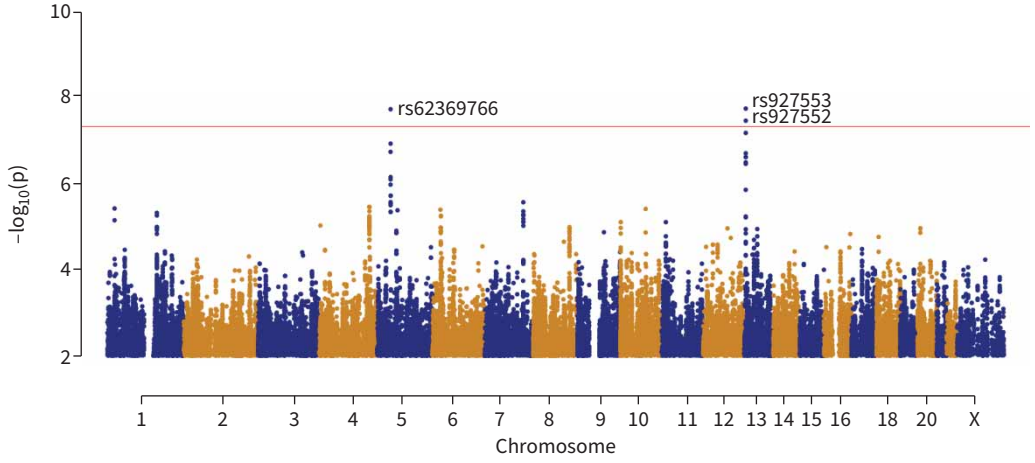


FIGURE 2 Manhattan plot for the genome-wide association study of chronic *Pseudomonas aeruginosa* age in 1037 individuals with cystic fibrosis. Each dot corresponds to a single nucleotide polymorphism (SNP) association p-value, with the *x*-axis indicating the position in the base-pair coordinate (hg19) and the *y*-axis indicating $-\log_{10}(p)$. In total, 6 994 213 SNPs were analysed, all with minor allele frequency >5%. The red horizontal line corresponds to the genome-wide significance threshold of $p=5\times10^{-8}$ on the $-\log_{10}$ scale. The QQ plot and histogram of the p-values are shown in supplementary figure S1, showing no inflation with a genomic control λ-value of 1.003.

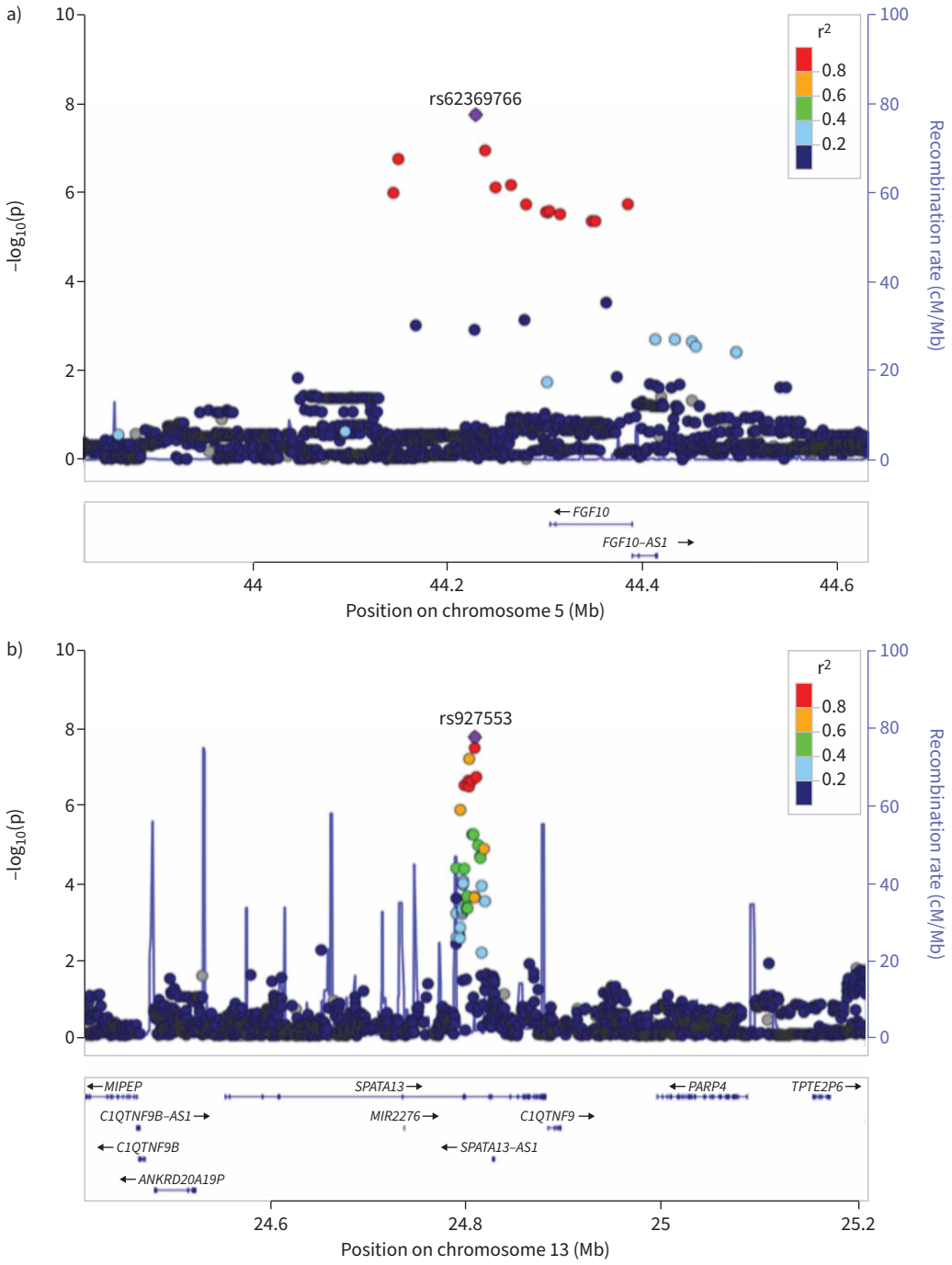


FIGURE 3 LocusZoom [63] region plots with the hg19 genome build for the two loci associated with the chronic *Pseudomonas aeruginosa* age residual: a) the chr5p12 locus tagged by rs62369766 and b) the chr13q12.12 locus tagged by rs927553. Each dot corresponds to a single nucleotide polymorphism (SNP) with the *x*-axis indicating the position in base-pair coordinates and the *y*-axis indicating $-\log_{10}(p)$. The colours represent the 1000 Genomes EUR linkage disequilibrium r^2-values with the top SNP (shown as the purple diamond and labelled with the dbSNP ID).

our study (table 3; each copy of coded allele T is associated with 2-year earlier chronic *P. aeruginosa* infection in our study and a decrease of 0.01–0.02 standard deviations in lung function in SHRINE *et al.* [30]; details on the spirometry measurements for their studies are available in their supplementary tables 2 and 3 and in their supplementary note).

TABLE 2 Three single nucleotide polymorphisms (SNPs) significantly associated with chronic *Pseudomonas aeruginosa* age at the genome-wide significance level

SNP	Chromosome	Effect allele	Reference allele	Sample	n	MAF	β[#] (95% CI) (years)	p-value
rs927553	13	G	C	CGMS	1037	0.44	1.51 (0.98–2.04)	1.91×10^{-8}
				European	991	0.43	1.53 (0.98–2.07)	3.33×10^{-8}
				Non-European	46	0.49	1.15 (−0.82–3.11)	0.25
				FCGS	501	0.45	0.05 (−0.68–0.77)	0.90
rs927552	13	G	A	CGMS	1037	0.41	1.51 (0.97–2.04)	3.64×10^{-8}
				European	991	0.40	1.55 (1.00–2.11)	3.88×10^{-8}
				Non-European	46	0.47	0.64 (−1.19–2.47)	0.49
				FCGS	501	0.41	0.01 (−0.71–0.73)	0.97
rs62369766	5	T	G	CGMS	1037	0.17	−2.00 (−2.70– −1.30)	1.98×10^{-8}
				European	991	0.17	−1.96 (−2.68– −1.25)	7.13×10^{-8}
				Non-European	46	0.10	−2.91 (−5.92–0.10)	0.06
				FCGS	501	0.17	−1.24 (−2.19– −0.30)	0.01

MAF: minor allele frequency. The genetic effects and p-values are provided, separately, for association analyses using 1) the combined sample of 1037 individuals in the Canadian CF Gene Modifier Study (CGMS), 2) 991 European individuals (including 52 sibling pairs) in CGMS, 3) 46 non-European individuals (including five sibling pairs) in CGMS and 4) 501 individuals in the independent French CF Modifier Gene Study (FCGS). The association analysis accounted for sample relatedness through the linear mixed effect model (supplementary appendix S1). [#]: effect estimate for each additional copy of the effect allele.

Colocalisation of CF lung disease and chronic P. aeruginosa *age GWAS summary statistics at lung disease-associated loci*

We examined our GWAS results at loci previously reported to be associated with chronic *P. aeruginosa* age and lung disease in CF. No previously reported SNPs [14, 24, 31] demonstrated significance at the level of 0.05 in our GWAS (supplementary table S2). Although none of the individual CF lung

TABLE 3 Supporting evidence for rs62369766 from the genome-wide meta-analysis of lung function in the non-cystic fibrosis (CF) population from Shrine *et al.* [30]

Coded allele	Non-coded allele	Trait	Study	Coded allele frequency	β[#]	SE	z-score	p-value
Chronic *P. aeruginosa* age GWAS								
T	G	Chronic *P. aeruginosa* age	CGMS (n=1037)	0.17	−2.00 (year)	0.36	−5.61	1.98E-8
Shrine *et al.* [30]								
T	G	FEV_1	UK Biobank (n=321 047)	0.16	−0.014	3.53E-3	−3.97	7.23E-5
			SpiroMeta (n=79 055)	0.15	−0.020	7.64E-3	−2.67	7.53E-3
			Meta-analysis (n=400 102)	0.16	−0.015	3.20E-3	−4.62	3.81E-6
		FVC	UK Biobank (n=321 047)	0.16	−0.014	3.55E-3	−4.04	5.46E-5
			SpiroMeta (n=79 055)	0.15	−0.020	7.66E-3	−2.58	9.82E-3
			Meta-analysis (n=400 102)	0.16	−0.015	3.20E-3	−4.70	2.59E-6
		PEF	UK Biobank (n=321 047)	0.16	−0.013	3.57E-3	−3.78	1.55E-4
			SpiroMeta (n=24 218)	0.16	−6.21E-3	0.012	−0.48	0.628
			Meta-analysis (n=345 265)	0.16	−0.013	3.00E-3	−3.69	2.25E-4
		FEV_1/FVC	UK Biobank (n=321 047)	0.16	−8.68E-4	3.62E-3	−0.33	0.740
			SpiroMeta (n=79 055)	0.15	−7.73E-4	7.58E-3	−0.10	0.919
			Meta-analysis (n=400 102)	0.16	−9.00E-4	3.30E-3	−0.26	0.794

Summary statistics for spirometric phenotypes are derived from a genome-wide meta-analysis of lung function in non-CF individuals of European descent, combining data from the UK Biobank (n=321 047) and the SpiroMeta consortium (n=79 055). This analysis, reported by Shrine *et al.* [30], included a total of 400 102 individuals. The phenotypes are ranked and inverse-normal transformed residuals from linear regression of each trait (forced expiratory volume in 1 s (FEV_1), forced vital capacity (FVC), peak expiratory flow (PEF) and FEV_1/FVC) against age, age squared, sex, height, smoking status (ever/never) and genotyping array. Individual study characteristics are summarised in supplementary tables 2 and 3 of Shrine *et al.* [30]. Summaries of the spirometry measurements for each study are available in their supplementary note. [#]: effect estimate for each additional copy of the effect allele T. GWAS: genome-wide association study; CGMS: Canadian CF Gene Modifier Study.

disease-associated SNPs were significant in our GWAS, we did find evidence at the locus level. That is, we found compelling evidence of colocalisation between GWAS summary statistics of CF lung disease and chronic *P. aeruginosa* age at SaKnorm-associated locus *AGTR2*/*SLC6A14* at chrXq22–q23 (supplementary table S3) [24]. In contrast, the genome-wide significant loci on chr5p12 and chr13q12.12 identified from the chronic *P. aeruginosa* age GWAS showed no evidence of colocalisation with the CF lung function GWAS summary statistics.

CF lung function PRS is significantly associated with the chronic P. aeruginosa *age residual*

We observed phenotypic correlation between chronic *P. aeruginosa* age and SaKnorm: Pearson correlation coefficient=0.12 (p=0.0002) in the unrelated European CGMS sample (n=887). Moreover, the identified genome-wide significant SNP rs62369766 for chronic *P. aeruginosa* age is also associated with respiratory phenotypes in the general population. These findings led us to consider the possibility of genetic overlap between chronic *P. aeruginosa* age and lung function in CF and non-CF populations. We subsequently investigated the potential genetic overlap through a cross-trait PRS analysis.

Using the C+T (clumping+thresholding) method (supplementary appendix S2), the PRS calculated from 7756 independent SNPs at an optimal threshold $p \leqslant 0.014$ (supplementary appendix S2 and supplementary table Sii) was associated with SaKnorm (p=0.002) (supplementary appendix S2 and supplementary figure Si) and explained 3.6% of the phenotypic variance in the chronic *P. aeruginosa* age residual. The effect size estimate was 0.618, indicating that a higher PRS score was associated with a later chronic *P. aeruginosa* age. The PRS analysis does not provide additional insight on particular contributing genes but rather that there are shared genetic mechanisms between lung function and chronic *P. aeruginosa* age (supplementary appendix S2 and supplementary table Si). Aligned with heritability estimates, there was no significant association between the PRS derived from lung function and first *P. aeruginosa* infection age (supplementary appendix S2 and supplementary figure Si).

Bidirectional causal relationship between chronic P. aeruginosa *infection and SaKnorm*

The significant genetic overlap between chronic *P. aeruginosa* age and lung function further led us to consider their causal relationship. Using bidirectional MR, we examined the causal link between chronic *P. aeruginosa* infection and lung disease in CF, utilising summary statistics from our GWAS (n=1037) and the CF lung function GWAS [24] (n=6365).

We first tested whether the earlier onset of chronic *P. aeruginosa* infection affects lung function in CF. A significant causal effect (β=0.002, p=0.011) of chronic *P. aeruginosa* infection on lung function was estimated using the inverse variance weighted (IVW) method. Due to the potential inclusion of a weak instrument (F-statistic=9.38), the causal effect estimate should be interpreted with caution. Further sensitivity analysis, based on MR methods accounting for bias arising from weak IVs, outliers and horizontal pleiotropy, did not support strong causal effects (p-values >0.05 for MR-Egger, MR weighted median-NOME and MR-PRESSO) (supplementary appendix S3 and supplementary table Si).

Reversely, we found evidence that reduced lung function leads to earlier chronic *P. aeruginosa* age, with significant causal effect estimate 0.78 ($p=4.23\times10^{-4}$ under the IVW method). Sensitivity analysis accounting for horizontal pleiotropy with MR-Egger (p=0.06) and the weighted median (p=0.01) show consistent effect estimates (supplementary appendix S3 and supplementary table Si). The causal effect was consistently observed with more stringent p-value thresholds for IV selection (supplementary appendix S3 and supplementary table Si).

In conclusion, our findings suggest a robust causal effect of diminished lung function on an earlier onset of chronic *P. aeruginosa* infection, with a moderate effect in the reverse direction. We note that the validity of these inferences is dependent on the quality and sample size of the GWAS data and the reliability of the instrumental variables used.

Discussion

The major cause of morbidity and mortality in CF is progressive lung disease, where cycles of infection and inflammation result in bronchiectasis and end-stage lung disease [32, 33]. Thus, an earlier acquisition of infection would be predicted to hasten this process. We set out to understand the genetic architecture of age of onset of *P. aeruginosa* acquisition: is it heritable, is it polygenic, are the genetic contributors unique to CF or do they contribute to lung function variation and infection phenotypes in the general population? SNP-sense heritability estimation [34] confirmed earlier reports from family studies [11] that age of chronic *P. aeruginosa* infection is more heritable than age of the first infection, focusing our choice of phenotype for the GWAS.

In the discovery GWAS using Canadian participants, we identified two common genetic variants, rs62369766 and rs927553, associated with the age of chronic *P. aeruginosa* infection residual in individuals with CF. The effect size and direction were consistent in the GWAS of European and non-European participants. These two SNPs explained ∼3% of variation in chronic *P. aeruginosa* infection age. The association at rs62369766 was validated using an independent French cohort. This analysis used a different phenotype and the analysis did not include older birth cohorts. According to our PheWAS results, the rs62369766 variant was associated with several spirometry phenotypes in non-CF populations [30], presenting robust evidence across several independent datasets.

Although a GWAS identifies loci rather than genes, the two loci encompass interesting candidate genes with previously reported impact on lung function and the respiratory system. According to the Genotype-Tissue Expression (GTEx) portal [35], the genome-wide significant SNP rs62369766 (chr5p12) is an expression quantitative trait locus (eQTL) for *NNT* (nicotinamide transhydrogenase, mitochondrial) across several tissues including lung, with $p=4.7\times10^{-3}$. *NNT* plays a pivotal role in sustaining cellular redox balance and is increasingly recognised as a participant in immune response regulation [36]. It has notably been shown to modulate the immune response and host defence against pathogens [37]. Overexpression of *NNT* in a macrophage cell line resulted in defective secretion of pro-inflammatory cytokines and inefficiency in the clearance of intracellular bacteria (*Escherichia coli*). Further, mice with deletion of the *NNT* gene are more resistant to acute pulmonary infection caused by *Streptococcus pneumoniae* [37]. A redox imbalance with higher oxidative stress is well known in CF and promotes inflammation [38]. As a modulator of oxidative stress, it would therefore be interesting to understand the role of *NNT* in the context of CF and its functional involvement in *P. aeruginosa* infection.

Although *NNT* is compelling as the responsible gene, rs62369766 is an intergenic SNP near *FGF10* (fibroblast growth factor 10), a driver of branching morphogenesis in the lung airway. Signalling defects of *FGF10* during development are reported to cause neonatal lung disease, while in adulthood *FGF10* mutations may lead to airway abnormalities which increase the risk of chronic airway disease [39]. Specifically, individuals with CF tend to have congenital abnormalities in the shape and size of the airway [39–41]. Moreover, infants with CF demonstrate airway abnormalities, such as smaller calibre tracheas, before developing clinically apparent respiratory bacterial infection including *P. aeruginosa* [39, 42]. *FGF10* is involved in the development of submucosal glands, which are important sources of host defence molecules [43, 44]. The association of SNP rs62369766 with reduced lung function is noteworthy considering that severe mutations in *FGF10* can disrupt lung development and may predispose individuals to childhood interstitial lung disease [24, 45].

Since the presence of chronic *P. aeruginosa* infection leads to progressive lung disease [4], we were interested in whether variants identified through a GWAS of CF lung function also contribute to an earlier age of chronic *P. aeruginosa* infection. Five loci display significant association with variation in CF lung disease [24]. Although none of the top SNPs at these loci, individually, showed association with chronic *P. aeruginosa* infection age at the 5% level in the GWAS reported here (supplementary table S3), we did see colocalisation at the locus level at SaKnorm-associated locus *AGTR2*/*SLC6A14* at chrXq22–q23. We also hypothesised that there is an overlap in CF genetic contributors between chronic *P. aeruginosa* age and CF lung disease at the sub-genome-wide significance level. The PRS defined on summary statistics from Corvol *et al.* [24] showed significant association with the chronic *P. aeruginosa* age residual, indicating that there is a shared polygenic component between the two phenotypes.

Our PRS analysis is primarily designed to identify the shared genetic factors between chronic *P. aeruginosa* infection and lung function, without specifically pinpointing if these factors are purely modifiers of severe lung disease or if they confer additional, distinct susceptibility to *P. aeruginosa* infection. Our findings suggest that some of the same genetic modifiers influencing severe lung disease may also affect the age of onset of chronic *P. aeruginosa*. This does not exclude the possibility that other genetic factors specifically enhance abnormal host defence mechanisms against *P. aeruginosa* infection, which the PRS analysis is not equipped to assess. Moreover, although the CF lung function PRS explained a significant proportion of variation in age of chronic *P. aeruginosa*, it would not have captured all the genetic factors given the limited sample size of the study. A limitation of this analysis, which may contribute to the low estimate of correlation between the two phenotypes, is the variability in timing of SaKnorm measurements relative to the chronic *P. aeruginosa* age.

The shared polygenic component between chronic *P. aeruginosa* age and CF lung function led us to consider their potential causal relationship. CF arises from genetic variation in *CFTR* that disrupts chloride transport *via* the apical membrane of epithelial cells. Loss of *CFTR* function weakens respiratory defences,

increasing susceptibility to bacterial infections, and leading to chronic inflammation and progressive lung disease in CF patients [32, 33, 46]. Two potential causal pathways are: 1) reduced lung function, stemming from genetic predisposition, could lead to earlier chronic *P. aeruginosa* age or 2) earlier chronic *P. aeruginosa* age might result in reduced lung function. Using bidirectional MR, we identified strong and robust evidence for a causal effect of reduced lung function on earlier chronic *P. aeruginosa* age ($p=4.23\times10^{-4}$) and a moderate causal effect for the direction of earlier chronic *P. aeruginosa* age on reduced lung function ($p=0.01$). Our findings corroborate the clinical perspective regarding the cyclic nature of escalating infection and diminishing lung function in CF [32, 33]. Moreover, the causal effect magnitude differences between the two directions may explain the previous results, that although earlier *P. aeruginosa* infection is strongly associated with severe CF lung disease later in life [16], it is a weaker predictor than early compromised childhood lung function [47]. While MR analysis provides evidence for a causal relationship between chronic *P. aeruginosa* infection and lung function, it does not precisely disentangle the timing of this causal effect due to the multifactorial nature of lung disease progression in CF.

There are many extrinsic factors that affect the age at chronic *P. aeruginosa* infection in CF. Certain *P. aeruginosa* strains are known to be more virulent and more transmissible among people with CF; exposure to these strains may shift the age of chronic infection to younger years [48]. Although the majority of people with CF are thought to acquire initial *P. aeruginosa* from environmental sources, patient-to-patient transmission of *P. aeruginosa* has been described, both in healthcare and non-healthcare settings [49]. Within households, siblings with CF frequently share *P. aeruginosa* strains due to cross-infection, limiting the generalisability of heritability estimates from the twin study by Green *et al.* [11] to the broader CF population [50]. Moreover, Green *et al.*'s [11] definition of the first *P. aeruginosa* infection requires an initial negative culture to accurately assign the first positive culture. This requirement may exclude participants with early infections, but it ensures precise assignment and prevents overestimating infection ages. For instance, individuals without prior negative cultures might have had an earlier unobserved infection than our data shows. Sensitivity analysis (supplementary appendix S6) confirmed consistent results when 565 individuals with censored event times were included. Accurately measuring initial and chronic infections in an observational cohort study is challenging, especially without extensive longitudinal records. We recognise that accurately measuring the first and chronic infection events is complex in an observational cohort study, particularly when dense longitudinal records are not available. Additionally, the institution of infection control practices over the last several decades has undoubtedly reduced *P. aeruginosa* transmission among people with CF and led to an increased age of chronic infection [51–56]. Differences in clinical management between CF centres as well as improvements in the quality of care may also have impacted the age of chronic *P. aeruginosa* infection over the course of our study [57]. Finally, different types of respiratory specimens (throat swab, induced sputum and bronchoalveolar lavage) [58, 59] and different diagnostic methods (culture *versus* molecular-based techniques) [60] have varying sensitivity in the detection of organisms, influencing the age of onset of chronic infection. The number of external elements that contribute to *P. aeruginosa* infection in people with CF may thus limit the ability to detect host genetic factors, accounting for the discovery of only two novel loci linked to chronic *P. aeruginosa* infection age. The GWAS sample size of n=1037 may have also limited the detection of significant loci that modify chronic *P. aeruginosa* age. Similar sample size challenges may have also affected previous studies on genetic modifiers of *P. aeruginosa* infection age (supplementary table S2) [14].

Here we identify two genome-wide significant loci for chronic *P. aeruginosa* age and support previous findings that the age of first *P. aeruginosa* acquisition is less heritable and potentially more opportunistic. eQTL support at one locus suggests that the responsible gene may be *NNT* and acting directly through host defence mechanisms, although the GWAS and eQTL statistics do not colocalise in general [61, 62]. Analysis of this SNP in the phenome of the general population indicates contributions to lung function, while PRS analysis demonstrates a shared polygenic component between chronic *P. aeruginosa* age and lung function in CF, and to a lesser degree, in the general population (supplementary appendix S2 and supplementary table Sii). There is also evidence for a causal effect of reduced lung function on the age of chronic *P. aeruginosa* infection and a moderate causal effect in the reverse direction, albeit less robust. Taken together, there is substantial overlap in the genetic contributions of earlier acquisition of chronic *P. aeruginosa* and lung disease in CF. Therapeutics and treatment modalities that aim to delay chronic infection should thus be a priority to preserve lung function.

Acknowledgements: We thank the CF participants, care providers and clinic coordinators at CF centres throughout Canada and France. We thank CF Canada for access to the Canadian CF Patient Registry and the Canadian CF Gene Modifier Study and the French CF Modifier Gene Study collaborators.

Data availability: All data analysed here are available by application to the Canadian CF Registry for researchers who meet the criteria for access to confidential clinical data for the purpose of CF research. Requests to access these datasets should be directed to cfregistry@cysticfibrosis.ca. Full summary statistics for the genome-wide meta-analysis of this study can be accessed from GitHub (https://github.com/strug-hub/gwas_psa). The summary statistics for the CF lung analysis are available at https://lab.research.sickkids.ca/strug/softwareandresources/ under the "Nature Communication 2015" tab under the "Resources" section. The summary statistics from the non-CF GWAS used in the paper are available from Shrine *et al.* [30] or available through application to the UK Biobank.

Ethics approval: We obtained written informed consent from adults, or from parents or guardians for participants aged <18 years old. The Canadian CF Gene Modifier Study was approved by the Hospital for Sick Children Research Ethics Board with research ethics approval number 1000065760. The French CF Modifier Gene Study was approved by the French ethical committees and written informed consent was obtained from each patient and/or guardian.

Conflict of interest: The authors have no potential conflicts of interest to disclose.

Support statement: Funding was provided by peer-reviewed Cystic Fibrosis (CF) Canada 2022 Basic and Clinical Research Grant (1009794) jointly funded by CF Canada and Canadian Institutes of Health Research Institute of Circulatory and Respiratory Health (CIHR-ICRH) FRN: BCG 187014 and Cystic Fibrosis Canada Grant (608828); the Cystic Fibrosis Foundation (STRUG17PO); Canadian Institutes of Health Research Foundation Grant (FRN-167282), and by the Government of Canada through Genome Canada and Ontario Genomics Institute (OGI-148). This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program to L.J. Strug who is the Canada Research Chair in Genome Data Science. The funders of the study played no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. French work was supported by Vaincre la Mucoviscidose, Blanche pour Vaincre la Mucoviscidose and Agir et Informer Contre la Mucoviscidose. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Gibson RL, Burns JL, Ramsey BW. Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am J Respir Crit Care Med* 2003; 168: 918–951.
2 Gong J, Wang F, Xiao B, *et al.* Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *PLoS Genet* 2019; 15: e1008007.
3 Blanchard AC, Waters VJ. Opportunistic pathogens in cystic fibrosis: epidemiology and pathogenesis of lung infection. *J Pediatric Infect Dis Soc* 2022; 11: Suppl. 2, S3–S12.
4 Kosorok MR, Zeng L, West SE, *et al.* Acceleration of lung disease in children with cystic fibrosis after *Pseudomonas aeruginosa* acquisition. *Pediatr Pulmonol* 2001; 32: 277–287.
5 McColley SA, Schechter MS, Morgan WJ, *et al.* Risk factors for mortality before age 18 years in cystic fibrosis. *Pediatr Pulmonol* 2017; 52: 909–915.
6 Mesinele J, Ruffin M, Kemgang A, *et al.* Risk factors for *Pseudomonas aeruginosa* airway infection and lung function decline in children with cystic fibrosis. *J Cyst Fibros* 2022; 21: 45–51.
7 Emerson J, Rosenfeld M, McNamara S, *et al. Pseudomonas aeruginosa* and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol* 2002; 34: 91–100.
8 Kiedrowski MR, Bomberger JM. Viral-bacterial co-infections in the cystic fibrosis respiratory tract. *Front Immunol* 2018; 9: 3067.
9 Cystic Fibrosis Canada. The Canadian Cystic Fibrosis Registry 2021 Annual Data Report. 2023. www.cysticfibrosis.ca/uploads/2021-Annual-Data-Report-WEB-AODA.pdf Date last accessed: 27 July 2024.
10 Cystic Fibrosis Foundation. Cystic Fibrosis Foundation Patient Registry 2021 Annual Data Report. 2022. www.cff.org/sites/default/files/2021-11/Patient-Registry-Annual-Data-Report.pdf Date last accessed: 27 July 2024.
11 Green DM, Collaco JM, McDougal KE, *et al.* Heritability of respiratory infection with *Pseudomonas aeruginosa* in cystic fibrosis. *J Pediatr* 2012; 161: 290–295.
12 Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet* 2015; 16: 45–56.
13 Rosenfeld M, Emerson J, McNamara S, *et al.* Risk factors for age at initial *Pseudomonas* acquisition in the cystic fibrosis epic observational cohort. *J Cyst Fibros* 2012; 11: 446–453.
14 Emond MJ, Louie T, Emerson J, *et al.* Exome sequencing of extreme phenotypes identifies *DCTN4* as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet* 2012; 44: 886–889.
15 Chalmers JD, Fleming GB, Hill AT, *et al.* Impact of mannose-binding lectin insufficiency on the course of cystic fibrosis: a review and meta-analysis. *Glycobiology* 2011; 21: 271–282.
16 Pittman JE, Calloway EH, Kiser M, *et al.* Age of *Pseudomonas aeruginosa* acquisition and subsequent severity of cystic fibrosis lung disease. *Pediatr Pulmonol* 2011; 46: 497–504.
17 Lin B, Gong J, Panjwani N, *et al.* 68 A genome-wide association study identifies two novel loci for respiratory infection with *Pseudomonas aeruginosa* in cystic fibrosis. *Genet Epidemiol* 2021; 45: 770.

18 Panjwani N, Xiao B, Xu L, *et al.* Improving imputation in disease-relevant regions: lessons from cystic fibrosis. *NPJ Genom Med* 2018; 3: 8.
19 McCaw ZR, Lane JM, Saxena R, *et al.* Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* 2020; 76: 1262–1272.
20 Milošević B, Obradović M. Comparison of efficiencies of some symmetry tests around an unknown centre. *Statistics* 2018; 53: 43–57.
21 Yang J, Lee SH, Goddard ME, *et al.* GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; 88: 76–82.
22 Gogarten SM, Sofer T, Chen H, *et al.* Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics* 2019; 35: 5346–5348.
23 Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008; 32: 227–234.
24 Corvol H, Blackman SM, Boelle PY, *et al.* Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nat Commun* 2015; 6: 8382.
25 Taylor C, Commander CW, Collaco JM, *et al.* A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. *Pediatr Pulmonol* 2011; 46: 857–869.
26 Hemani G, Zheng J, Elsworth B, *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; 7: e34408.
27 Bowden J, Davey Smith G, Haycock PC, *et al.* Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304–314.
28 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.
29 Watanabe K, Stringer S, Frei O, *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019; 51: 1339–1348.
30 Shrine N, Guyatt AL, Erzurumluoglu AM, *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019; 51: 481–493.
31 Jackson VE, Latourelle JC, Wain LV, *et al.* Meta-analysis of exome array data identifies six novel genetic loci for lung function. *Wellcome Open Res* 2018; 3: 4.
32 Nichols D, Chmiel J, Berger M. Chronic inflammation in the cystic fibrosis lung: alterations in inter- and intracellular signaling. *Clin Rev Allergy Immunol* 2008; 34: 146–162.
33 Cohen-Cymberknoh M, Kerem E, Ferkol T, *et al.* Airway inflammation in cystic fibrosis: molecular mechanisms and clinical implications. *Thorax* 2013; 68: 1157–1162.
34 Yang J, Zeng J, Goddard ME, *et al.* Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet* 2017; 49: 1304–1310.
35 Lonsdale J, Thomas J, Salvatore M, *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; 45: 580–585.
36 Regan T, Conway R, Bharath LP. Regulation of immune cell function by nicotinamide nucleotide transhydrogenase. *Am J Physiol Cell Physiol* 2022; 322: C666–C673.
37 Ripoll VM, Meadows NA, Bangert M, *et al.* Nicotinamide nucleotide transhydrogenase (NNT) acts as a novel modulator of macrophage inflammatory responses. *FASEB J* 2012; 26: 3550–3562.
38 Moliteo E, Sciacca M, Palmeri A, *et al.* Cystic fibrosis and oxidative stress: the role of CFTR. *Molecules* 2022; 27: 5324.
39 Prince LS. FGF10 and human lung disease across the life spectrum. *Front Genet* 2018; 9: 517.
40 Fischer AJ, Singh SB, Adam RJ, *et al.* Tracheomalacia is associated with lower FEV_1 and *Pseudomonas* acquisition in children with CF. *Pediatr Pulmonol* 2014; 49: 960–970.
41 Meyerholz DK, Stoltz DA, Namati E, *et al.* Loss of cystic fibrosis transmembrane conductance regulator function produces abnormalities in tracheal development in neonatal pigs and young children. *Am J Respir Crit Care Med* 2010; 182: 1251–1261.
42 Sly PD, Brennan S, Gangell C, *et al.* Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *Am J Respir Crit Care Med* 2009; 180: 146–152.
43 Rawlins EL, Hogan BL. Intercellular growth factor signaling and the development of mouse tracheal submucosal glands. *Dev Dyn* 2005; 233: 1378–1385.
44 Widdicombe JH, Wine JJ. Airway gland structure and function. *Physiol Rev* 2015; 95: 1241–1319.
45 Karolak JA, Vincent M, Deutsch G, *et al.* Complex compound inheritance of lethal lung developmental disorders due to disruption of the TBX–FGF pathway. *Am J Hum Genet* 2019; 104: 213–228.
46 Lei L, Traore S, Romano Ibarra GS, *et al.* CFTR-rich ionocytes mediate chloride absorption across airway epithelia. *J Clin Invest* 2023; 133: e171268.
47 Pittman JE, Noah H, Calloway HE, *et al.* Early childhood lung function is a stronger predictor of adolescent lung function in cystic fibrosis than early *Pseudomonas aeruginosa* infection. *PLoS One* 2017; 12: e0177215.

48 Parkins MD, Somayaji R, Waters VJ. Epidemiology, biology, and impact of clonal *Pseudomonas aeruginosa* infections in cystic fibrosis. *Clin Microbiol Rev* 2018; 31: e00019-18.

49 Stapleton PJ, Izydorcyzk C, Clark S, *et al. Pseudomonas aeruginosa* strain-sharing in early infection among children with cystic fibrosis. *Clin Infect Dis* 2021; 73: e2521–e2528.

50 Wiehlmann L, Wagner G, Cramer N, *et al.* Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 2007; 104: 8101–8106.

51 Saiman L. Infection prevention and control in cystic fibrosis. *Curr Opin Infect Dis* 2011; 24: 390–395.

52 Saiman L, Garber E. Infection control in cystic fibrosis: barriers to implementation and ideas for improvement. *Curr Opin Pulm Med* 2009; 15: 626–631.

53 Zhou J, Garber E, Saiman L. Survey of infection control policies for patients with cystic fibrosis in the United States. *Am J Infect Control* 2008; 36: 220–222.

54 Saiman L, Siegel J. Infection control in cystic fibrosis. *Clin Microbiol Rev* 2004; 17: 57–71.

55 Saiman L, Siegel J, Cystic Fibrosis Foundation Consensus Conference on Infection Control Participants. Infection control recommendations for patients with cystic fibrosis: microbiology, important pathogens, and infection control practices to prevent patient-to-patient transmission. *Am J Infect Control* 2003; 31: Suppl. 3, S1–S62.

56 Saiman L, Siegel JD, LiPuma JJ, *et al.* Infection prevention and control guideline for cystic fibrosis: 2013 update. *Infect Control Hosp Epidemiol* 2014; 35: Suppl. 1, S1–S67.

57 Shah KS, Saiman L, LiPuma JJ, *et al.* Association of *Pseudomonas aeruginosa* incident infections with adherence to Cystic Fibrosis Foundation care guidelines. *J Cyst Fibros* 2024; 23: 300–305.

58 Ronchetti K, Tame JD, Paisey C, *et al.* The CF-Sputum Induction Trial (CF-SpIT) to assess lower airway bacterial sampling in young children with cystic fibrosis: a prospective internally controlled interventional trial. *Lancet Respir Med* 2018; 6: 461–471.

59 Weiser R, Oakley J, Ronchetti K, *et al.* The lung microbiota in children with cystic fibrosis captured by induced sputum sampling. *J Cyst Fibros* 2022; 21: 1006–1012.

60 Nichols DP, Morgan SJ, Skalland M, *et al.* Pharmacologic improvement of CFTR function rapidly decreases sputum pathogen density, but lung infections generally persist. *J Clin Invest* 2023; 133: e167957.

61 Panjwani N, Wang F, Mastromatteo S, *et al.* LocusFocus: web-based colocalization for the annotation and functional follow-up of GWAS. *PLoS Comput Biol* 2020; 16: e1008336.

62 Wang F, Panjwani N, Wang C, *et al.* A flexible summary statistics-based colocalization method with application to the mucin cystic fibrosis lung disease modifier locus. *Am J Hum Genet* 2022; 109: 253–269.

63 Pruim RJ, Welch RP, Sanna S, *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; 26: 2336–2337.