SERPINA1 Z allele is associated with cystic fibrosis liver disease

Pierre-Yves Boëlle, Dominique Debray, Loïc Guillot, Harriet Corvol

SERPINA1 Z allele is associated with cystic fibrosis liver disease. Boëlle PY, Debray D, Guillot L, **Corvol H**; French CF Modifier Gene Study Investigators. Genet Med. 2019 Feb 11. doi: 10.1038/s41436-019-0449-6.

https://doi.org/10.1038/s41436-019-0449-6

SERPINA1 Z allele is associated with cystic fibrosis liver disease

Pierre-Yves Boëlle, PhD[1], Dominique Debray, MD, PhD[2,3], Loic Guillot, PhD[2*], and Harriet Corvol, MD, PhD[2,4*] on behalf of the French CF Modifier Gene Study Investigators

[1]Sorbonne Université, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique (iPLESP), AP-HP, Hôpital Saint-Antoine, Paris F-75012, France

[2]Sorbonne Université, INSERM, Centre de Recherche Saint-Antoine, CRSA, Paris F-75012, France

[3]AP-HP, Hôpital Necker Enfants Malades, Unité d'Hépatologie Pédiatrique, Paris F-75007, France

[4]AP-HP, Hôpital Trousseau, Service de Pneumologie Pédiatrique, Paris F-75012, France

*co-last authors

Corresponding Author:

Harriet Corvol

Centre de Recherche Saint-Antoine, Bâtiment Kourilsky

34 Rue Crozatier, Paris 75012, France.

Tel: (33) 1 44 73 61 74

Email: harriet.corvol@aphp.fr

Conflict of interest notification

The authors certify that they have no potential conflict of interest to declare regarding the subject matter or materials discussed in this manuscript.

Financial Support

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), Sorbonne Université, Assistance Publique-Hôpitaux de Paris (APHP), Agence Nationale de la Recherche (ANR), DGS, Association Vaincre La Mucoviscidose, Cystic Fibrosis Foundation (CORVOL16G0), Association Agir Informer Contre la Mucoviscidose, GIS-Institut des Maladies Rares, and Société Française de la Mucoviscidose.

Abstract

Purpose: The *SERPINA1* Z allele is associated with cystic fibrosis (CF)-related liver disease (CFLD), a common manifestation in patients with CF. We estimated CFLD incidence based on the *SERPINA1* genotype in 3,328 CF patients with CFLD-phenotype information.

Methods: The associations of *SERPINA1* Z (rs28929474) and S (rs17580) alleles with age at CFLD onset and the development of CFLD-related complications (severe liver disease with cirrhosis, portal hypertension, esophageal varices) were analyzed.

Results: Overall, 3% of patients carried the *SERPINA1* Z allele and 13% the S allele. The cumulative incidence of CFLD increased more rapidly in patients carrying the Z allele (HR = 1.6; 95% CI = 1.1–2.4, P = 0.019), reaching 47% by age 25 compared to 30% in non-carriers. Increased risk was similar for patients with severe CFLD (HR = 1.5, 95% CI = 0.7–3.2, P = 0.31) but failed to reach significance due to a limited sample size of Z-allele carriers. No significant effect was found for the S allele.

Conclusion: CF patients carrying the *SERPINA1* Z allele had an increased risk of developing CFLD and related complications compared to non-carriers. Routine *SERPINA1* Z genotyping upon CF diagnosis is warranted for identifying patients worthy of closer liver disease monitoring.

Keywords: SERPINA1 Z, cystic fibrosis, liver disease, cirrhosis, portal hypertension

INTRODUCTION

Cystic fibrosis (CF) is the most common, severe, autosomal recessive genetic disease in Caucasians. Pathogenic variants in the gene encoding the chloride channel CF transmembrane conductance regulator (CFTR)[1] induce multi-organ dysfunction, including the lungs, pancreas, intestine, and liver. A broad spectrum of hepatobiliary abnormalities are covered under the term CF-related liver disease (CFLD). Focal biliary cirrhosis is the most clinically important form of CFLD, since extension of the initially focal fibrogenic process may cause multilobular biliary cirrhosis followed by portal hypertension and associated complications.[2,3] Currently, multilobular cirrhosis ranks as the third leading cause of death in patients with CF after respiratory failure and transplantation-related complications.[4]

Although CF is recognized as a monogenic disease, considerable interindividual variability in phenotype exists among patients with the same *CFTR* pathogenic variants.[5-7] In addition to environmental factors, genetic modifiers have also been shown to contribute to this variability.[5,6] We have assembled a large cohort of patients as part of the French CF Modifier Gene Study and recently reported the incidence of CFLD and severe CFLD in these patients.[8] We found that the risk of CFLD increases with age, with a frequency up to 32% by age 25, and is associated with several risk factors, including male sex, *CFTR* F508del homozygosity, and a history of meconium ileus at birth.

The *SERPINA1* gene has been implicated in the development and progression of CFLD.[9] *SERPINA1* encodes the alpha-1 antitrypsin (AAT) protein synthesized by the liver. Several pathogenic variants of *SERPINA1* have been determined to cause an AAT deficiency that predisposes an individual to liver disease and early-onset emphysema. The most common variants involve the Z and S alleles, each caused by single nucleotide polymorphisms.[10] The Z variant is the allele overwhelmingly associated with liver disease.[11] Indeed, during biogenesis the Z-type AAT protein folds abnormally in the endoplasmic reticulum of hepatocytes and is retained intracellularly instead of being efficiently secreted, resulting in low serum levels of AAT.[12] The intracellular accumulation of AAT mutant Z proteins within hepatocytes can lead to liver injury, cirrhosis, or hepatocellular

carcinoma.[13] Individuals heterozygous for AAT that carry one normal protease inhibitor M allele and one pathogenic variant Z allele (PiMZ or MZ) are asymptomatic with regard to liver disease. Compound heterozygotes for the S and the Z alleles of AAT (PiSZ) may develop liver disease with identical manifestations to those of PiZZ patients, whereas liver disease is absent in PiSS homozygote individuals.[13] A CF case-control study revealed that carriers of the *SERPINA1* Z allele are more common among patients with severe CFLD, with a large odds ratio of 5.[9] However, the study focused on patients with severe liver disease defined as cirrhosis showing signs of portal hypertension, which prevented the determination of the actual risk associated with the variant for developing CFLD. Therefore, in the present study, we estimated the cumulative incidence of CFLD based on the *SERPINA1* genotype in a cohort of French CF patients of unprecedented size (n = 3,328).

MATERIAL AND METHODS

As previously described, we assembled the French CF Modifier Gene Study cohort that included CF patients treated at French CF centers since 2004 (see also **Supplementary Information**).[8] In brief, 4,798 patients with CF were recruited for the study, corresponding to approximately 80% of all French patients with CF.[14] Among this cohort, 3,328 CF patients with pancreatic insufficiency born after 1985 were available for evaluation of CFLD incidence and risk factors, including patients with severe CFLD.[8] CFLD was defined according to the European Best Practice Guidance by Debray *et al.*[2] Patients with cirrhosis, portal hypertension, and/or esophageal varices were classified as having severe CFLD (see also **Supplementary Information**).[3,15] The study was approved by the French ethical committee (CPP no. 2004/15) and the information collection was approved by the Commission Nationale de L'informatique et des Libertés (no. 04.404). Written informed consent was obtained from each patient and/or guardian.

Genotyping of the *SERPINA1* Z (rs28929474) and S (rs17580) alleles was carried out using Kompetitive Allele Specific PCR (KASP) genotyping chemistry (LGC, Teddington, UK). In the dbSNP database (https://www.ncbi.nlm.nih.gov/projects/SNP/), rs28929474 is identified as an A/G variant

with G being the ancestral allele, and rs17580 is identified as an A/T variant with A being the ancestral allele.

Descriptive statistics were compiled as the mean ± standard deviation (SD) or percentages as appropriate. All patients were considered to be at risk for CFLD since birth and were censored at the time of the last visit without a CFLD diagnosis before January 2017. For other patients, the age at CFLD diagnosis was determined using the date of the first report in the medical records, allowing for "interval censoring" between birth and the age of the first report when the date of onset was not precisely known. This was the case for 142 of 605 patients with CFLD (24%). Likewise, severe CFLD onset was defined as the first date that cirrhosis, portal hypertension, and/or esophageal varices were reported and uncertainty regarding this date was considered as described above. The date of severe CFLD was interval censored in 19 of 175 patients (11%). We used the log-rank test adapted for interval-censored data for comparisons between the cumulative incidences curves[16] and Cox regression adapted to interval-censored data to determine the association of factors linked to age at CFLD onset.[17] Confidence intervals were computed by the bootstrap method. Bonferroni correction was used for multiple comparisons.

RESULTS

Clinical characteristics of the 3,328 CF patients included in this study, along with the distribution of the *SERPINA1* Z and S alleles, are shown in **Table 1**. The minor allele frequencies were similar in our cohort to those reported for Europeans: *SERPINA1* Z (A variant): 1.4% vs. 2% respectively, *SERPINA1* S (T variant): 6.4% versus 6% respectively. There were no patients homozygous for *SERPINA1* Z. Details on the number of CF patients at risk of developing CFLD and severe CFLD, and the cumulative number of CFLD and severe CFLD events for the entire cohort as well as according to *SERPINA1* Z and S genotypes are provided in the **Supplementary Information (Table S1, S2 and S3).**

Overall, 3% of the CF patients carried the *SERPINA1* Z allele and 13% the S allele. The cumulative incidence of CFLD increased more rapidly in patients carrying the *SERPINA1* Z allele (HR = 1.6; 95% CI = 1.1–2.4, P = 0.019), reaching 47% by age 25 in the Z allele carriers compared to 30% at age 25 in the others (**Table 1** and **Figure 1**). The increase in risk was similar for patients with severe CFLD (HR = 1.5, 95% CI = 0.7–3.2, P = 0.31), but this did not reach statistical significance. With respect to severe CFLD, there were only seven cases among the Z allele carriers, making the cumulative incidence curve difficult to accurately estimate. Clinical characteristics such as sex, year of birth and of CF diagnosis were not associated with *SERPINA1* Z and S alleles (**Supplementary Information Table S4**). Adjusting on European origin, *CFTR* variants and meconium ileus did not change the strength of the association (**Supplementary Information Table S5**). The effect of carrying one *SERPINA1* S allele on CFLD risk was not statistically significant (**Table 1** and **Figure 1**).

DISCUSSION

The French CF Modifier Gene Study provided an unprecedentedly large cohort of 3,328 pancreatic-insufficient patients with CF born after 1985. This gave an opportunity to obtain a more accurate estimate of the incidence of CFLD and severe CFLD with sufficient power to detect associations of clinical relevance.[8] We found that the *SERPINA1* Z allele was associated with an increased risk of developing CFLD, although the association was weaker than that previously reported.[9] Nevertheless, the incidence of CFLD increased more rapidly in patients carrying the *SERPINA1* Z allele, with up to 47% of the Z allele carriers developing liver disease before the age of 25 compared to only 30% for non-carrier patients.

The role of *SERPINA1* in CFLD was first identified in a two-stage case-control study including CF patients from several countries worldwide.[9] Both the initial and a replicate studies showed that severe CFLD is associated with the *SERPINA1* Z allele (odds ratio of 4.72 and 3.42, respectively).[9] Analysis of our cohort of patients with CF confirmed this association, but demonstrated a smaller difference in risk than that reported previously. For example, using the cumulative incidence at age

25, the odds ratio for the Z allele was only 2 in our cohort. However, the difference in cumulative risk with age was still clinically relevant since *SERPINA1* Z carriers had a 50% greater risk of developing CFLD compared to non-carriers. We did not observe any association of the *SERPINA1* S allele with CFLD, which was not surprising since the S allele is recognized to be associated with reduced levels of AAT protein, but not with liver manifestations.[18]

Our study had limitations related to its design and the use of medical records as a primary source of information. However, we previously reported that selection bias due to differential mortality was likely to be small.[8] A second potential issue was the rarity of the *SERPINA1* Z allele in our cohort. This reduced the precision of estimation, despite the large cohort size. Indeed, there were only 90 carriers of the *SERPINA1* Z allele (PiMZ) and no homozygous PiZZ patients in the cohort. Furthermore, only seven of these patients experienced severe CFLD, making the cumulative incidence difficult to accurately estimate.

Obtaining a more accurate prediction of the risk of CFLD in CF patients remains an important issue, especially since ursodeoxycholic acid, the treatment commonly prescribed for its prevention, seems to have little to no effect.[8,19] Given this situation, identifying biomarkers to predict the occurrence of CFLD is fundamental to improving the monitoring of disease progression and for assessing the effects of novel therapies, such as bile-salt analogs, anti-fibrotics, and CFTR correctors and potentiators. *SERPINA1* Z genotyping at the time of CF diagnosis may help to single out a population of patients who should be more closely screened for liver disease. Gaining a better understanding of the genetic profiles of patients with CFLD will undoubtedly open new therapeutic avenues and help to develop prospective therapies that focus on high-risk groups.

ACKNOWLEDGEMENTS

The authors would like to thank the clinical research assistants involved in the data collection, namely, Pauline Touche, Tania Kombi, and Julie Mésinèle. We would like to especially thank the

patients and families who participated in this study and the contributing French CF Centers (see the list of investigators in the **Supplementary Information**).

References

1. Riordan JR, Rommens JM, Kerem B, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science.* Sep 08 1989;245(4922):1066-1073.
2. Debray D, Kelly D, Houwen R, Strandvik B, Colombo C. Best practice guidance for the diagnosis and management of cystic fibrosis-associated liver disease. *J Cyst Fibros.* Jun 2011;10 Suppl 2:S29-36.
3. Stonebraker JR, Ooi CY, Pace RG, et al. Features of Severe Liver Disease With Portal Hypertension in Patients With Cystic Fibrosis. *Clin Gastroenterol Hepatol.* Apr 5 2016.
4. Rowland M, Gallagher C, Gallagher CG, et al. Outcome in patients with cystic fibrosis liver disease. *J Cyst Fibros.* Jan 2015;14(1):120-126.
5. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. *Nat Rev Genet.* Jan 2015;16(1):45-56.
6. Guillot L, Beucher J, Tabary O, Le Rouzic P, Clement A, Corvol H. Lung disease modifier genes in cystic fibrosis. *The international journal of biochemistry & cell biology.* Jul 2014;52:83-93.
7. McKone EF, Emerson SS, Edwards KL, Aitken ML. Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study. *Lancet.* May 17 2003;361(9370):1671-1676.
8. Boelle PY, Debray D, Guillot L, Clement A, Corvol H, French CFMGSI. Cystic Fibrosis Liver Disease: Outcomes and Risk Factors in a Large Cohort of French Patients. *Hepatology.* Jul 30 2018.
9. Bartlett JR, Friedman KJ, Ling SC, et al. Genetic modifiers of liver disease in cystic fibrosis. *JAMA.* Sep 9 2009;302(10):1076-1083.

10. Bartels CL, Marchetti AL, Edward Highsmith W, Tsongalis GJ. Real time PCR detection of the PI*Z and PI*S mutations associated with alpha-1 antitrypsin deficiency. *American journal of translational research.* Aug 10 2009;1(4):406-411.

11. Zorzetto M, Russi E, Senn O, et al. SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. *Clinical chemistry.* Aug 2008;54(8):1331-1338.

12. Stoller JK, Aboussouan LS. A review of alpha1-antitrypsin deficiency. *Am J Respir Crit Care Med.* Feb 1 2012;185(3):246-259.

13. Teckman JH. Liver disease in alpha-1 antitrypsin deficiency: current understanding and future therapy. *Copd.* Mar 2013;10 Suppl 1:35-43.

14. Vaincre-la-Mucoviscidose, Ined. Registre Français de la mucoviscidose - Bilan des données 2016. 2017.

15. Debray D, Narkewicz MR, Bodewes F, et al. Cystic Fibrosis-related Liver Disease: Research Challenges and Future Perspectives. *J Pediatr Gastroenterol Nutr.* Oct 2017;65(4):443-448.

16. Fay MP. Rank invariant tests for interval censored data under the grouped continuous model. *Biometrics.* Sep 1996;52(3):811-822.

17. Pan W. Extending the Iterative Convex Minorant Algorithm to the Cox Model for Interval-Censored Data. *Journal of Computational & Graphical Statistics.* 1999;8:109--120.

18. Torres-Duran M, Lopez-Campos JL, Barrecheguren M, et al. Alpha-1 antitrypsin deficiency: outstanding questions and future directions. *Orphanet J Rare Dis.* Jul 11 2018;13(1):114.

19. Cheng K, Ashby D, Smyth RL. Ursodeoxycholic acid for cystic fibrosis-related liver disease. *Cochrane Database Syst Rev.* Sep 11 2017;9:CD000222.

Figure legend

Figure 1. Cumulative incidence of cystic fibrosis-related liver disease (CFLD) (A and C) and severe CFLD (B and D) according to *SERPINA1* Z (A and B) and S (C and D) alleles. In each graph, the solid-line curve indicates the cumulative incidence for patients carrying the normal M allele (G and A for *SERPINA1* Z and S, respectively) and the dotted-line curve indicates the cumulative incidence for patients carrying the pathogenic variant (A and T for *SERPINA1* Z and S, respectively). Below each graph is indicated the number of patients at a specific age who continue to be followed-up but have not yet develop CFLD or severe CFLD according to *SERPINA1* Z and S alleles.

Figure. 1

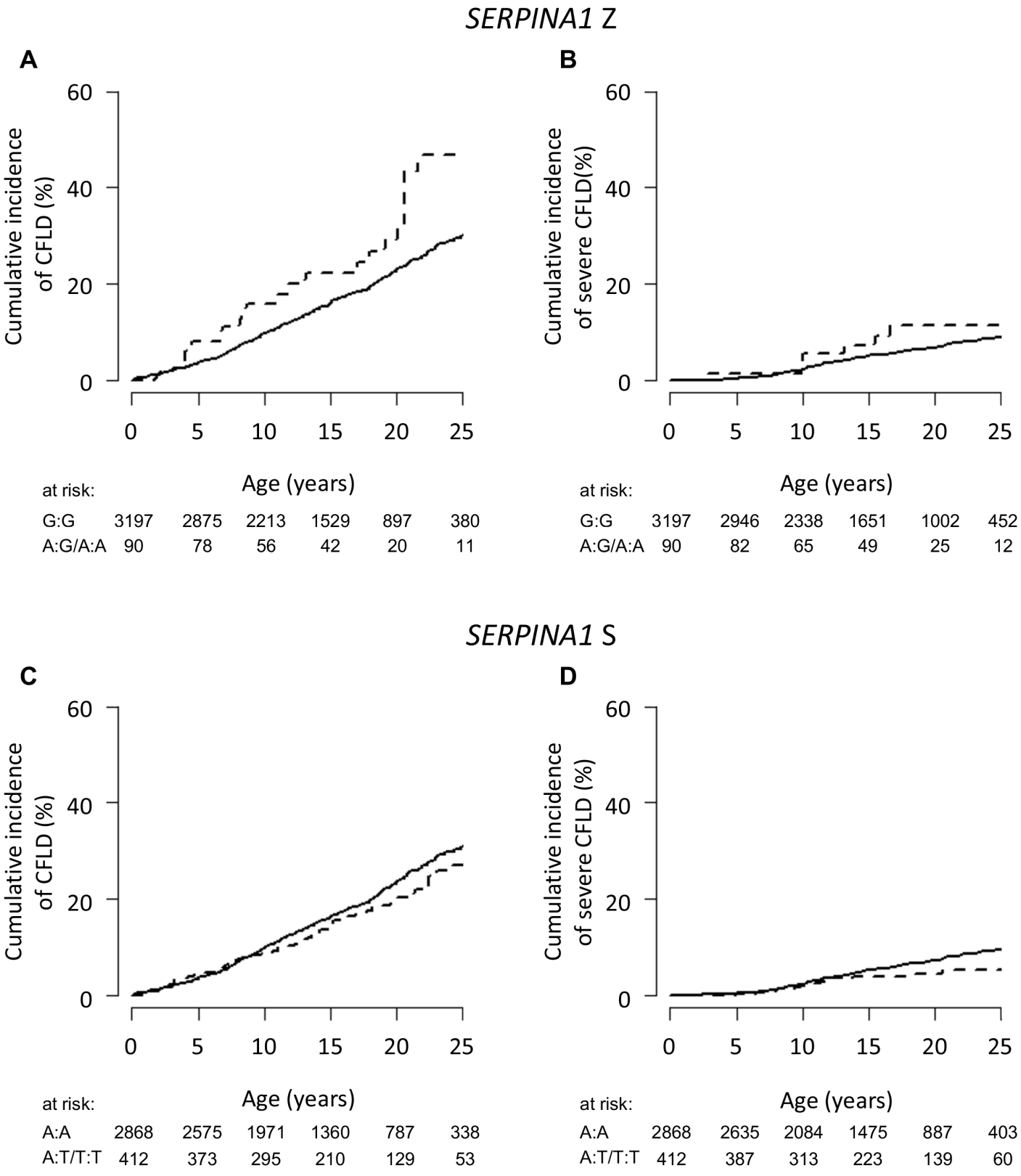
SERPINA1 Z
A
B
Cumulative incidence of CFLD (%)
Cumulative incidence of severe CFLD(%)
60
40
20
0
0 5 10 15 20 25
Age (years)
at risk:
G:G 3197 2875 2213 1529 897 380
A:G/A:A 90 78 56 42 20 11
at risk:
G:G 3197 2946 2338 1651 1002 452
A:G/A:A 90 82 65 49 25 12
SERPINA1 S
C
D
Cumulative incidence of CFLD (%)
Cumulative incidence of severe CFLD (%)
at risk:
A:A 2868 2575 1971 1360 787 338
A:T/T:T 412 373 295 210 129 53
at risk:
A:A 2868 2635 2084 1475 887 403
A:T/T:T 412 387 313 223 139 60

Table 1: Patient clinical characteristics, *SERPINA1* Z and S alleles distribution and their association with CFLD and severe CFLD.

Clinical characteristics	**Total (n)**	
Sex (Male), % (n)	3328	52.0% (1731)
Current age (years), mean ± SD		15.9 ± 7.7
Year of birth, % (n)		
1986-1995		35.5% (1183)
1996-2005		40.1% (1333)
2006-2017		24.4% (812)
Age at enrolment, % (n)		
Birth		35.2% (1172)
1-10 years		35.9% (1195)
10-18 years		28.9% (961)
European origin, % (n)		88.9% (2957)
CF diagnosis <1 year old, % (n)		73.9% (2461)
CFTR F508del homozygotes, % (n)		49.4% (1645)
Meconium ileus, % (n)		13.4% (445)
***SERPINA1* Z and S alleles distribution**	Total (n)	% (n)
***SERPINA1* Z**, % (n)	3287*	
G:G		97.3% (3197)
G:A		2.7% (90)
***SERPINA1* S**, % (n)	3280*	
A:A		87.4% (2868)
T:A		12.1% (396)
T:T		0.5% (16)
Association of *SERPINA1* Z and S alleles with CFLD and severe CFLD	**CFLD** HR (95%CI), P-value (Adjusted P-value**)	**Severe CFLD** HR (95%CI), P-value (Adjusted P-value**)
***SERPINA1* Z** (G:A vs G:G)	1.62 (1.08-2.42), P=0.019 (P_{adj}=0.076)	1.48 (0.70–3.14), P=0.31 (P_{adj}=1.0)
***SERPINA1* S** (T:A/T:T vs A/A)	0.87 (0.67-1.12), P=0.27 (P_{adj}=1.0)	0.64 (0.38-1.09), P=0.1 (P_{adj}=0.4)

Data are means ± SD or % (numbers) unless otherwise indicated. *Genotyping failed for 41 patients for the Z allele and 48 patients for the S allele. **Bonferroni adjustment for 4 tests. *Abbreviations*: CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator, CFLD: cystic fibrosis-related liver disease; HR: hazard ratio; 95% CI: 95% confidence interval.

SUPPLEMENTARY INFORMATION

Contents

List of French CF modifier Gene Study Investigators

ABELY,M., American Memorial Hospital, Reims / BASSINET,L., Centre Hospitalier Intercommunal de Créteil, Créteil / BELLEGUIC,C., Hôpital Pontchaillou, Rennes / BESSACI,K., American Memorial Hospital, Reims / BILLON,Y., Hôpital de Brabois, Vandoeuvre les Nancy / BRÉMONT,F., Hôpital des Enfants de Toulouse, Toulouse / BROUARD,J., Centre Hospitalier Universitaire de Caen, Caen / BUI,S., Hôpital Des Enfants Groupe Pellegrin, Bordeaux / CAMARA,B., Hôpital de la Tronche, Grenoble / CAMPBELL,K., Centre Hospitalier Universitaire de Caen, Caen / CHIRON,R., Hôpital Arnaud de Villeneuve, Montpellier / COIRIER-DUET,E., Hôpital André Mignot, Le Chesnay / DALPHIN,J.C., CNRS-UFC,UMR 6249 Chrono-environnement, Hôpital Jean Minjoz, Besançon /DALPHIN,M.L., Centre Hospitalier Universitaire de Besançon, Besançon / DAVID,V., Hôpital Mère-Enfant, Nantes / DENEUVILLE,E., Hôpital Sud Annexe Pédiatrique, Rennes / DERELLE,J., Hôpital d'Enfants, Vandoeuvre les Nancy / DOMBLIDES,P., Hôpital Haut Lévêque, Pessac / DOMINIQUE,S., Centre Hospitalier Universitaire Charles Nicolle, Rouen / DUBUS,J.C., Hôpital d'Enfants de la Timone, Marseille / DUFEU,N., Hôpital Nord, Marseille / DURIEU,I., UCBL1, Groupe Hospitalier Lyon Sud, Hospices Civils de Lyon, Pierre Bénite / DURY,S., Hôpital Maison Blanche, Reims / EPAUD,R., Centre Hospitalier Intercommunal de Créteil, Créteil / FANTON,A., Hôpital d'Enfants du Bocage, Dijon / FAYON,M., Hôpital Des Enfants Groupe Pellegrin, Bordeaux / FLEURENCE,E., Hôpital d'Enfants, Saint-Denis de la Réunion / FOUCAUD,P., Hôpital André Mignot, Le Chesnay / GINIES,J.L., Centre Hospitalier Universitaire d'Angers, Angers / GRENET,D., Hôpital Foch, Suresnes / HUBERT,D., Hôpital Cochin, Paris / HUET,F., Hôpital d'Enfants du Bocage, Dijon / KESSLER,R., Hôpital Civil, Strasbourg / LABBÉ,A., Centre Hospitalier Estaing, Clermont-Ferrand / LANGUEPIN,J., Centre Hospitalier, Limoges / LE BOURGEOIS,M., Necker Hôpital d'Enfants Malades, Paris / LE ROUX,P., Hôpital Jacques Monod, Montivilliers / LEROY,S., , CHU de Nice, Nice / LLERENA,C., Hôpital de la Tronche, Grenoble / LOEUILLE,G.A., Centre Hospitalier de Dunkerque, Dunkerque / MARCHAND,S., Hôpital Clocheville, Tours / MARGUET,C., Centre Hospitalier Universitaire Charles Nicolle, Rouen / MELY,L., Hôpital Renée Sabran, Giens / MUNCK,A., Hôpital Robert Debré, Paris / MURRIS-ESPIN,M., Hôpital Larrey, Toulouse

/ NOVE JOSSERAND,R., Groupe Hospitalier Lyon Sud, Hospices Civils de Lyon, Pierre Bénite / PIN,I., INSERM U823 Université Joseph Fourier, Hôpital de la Tronche, Grenoble / PREVOTAT,A., Hôpital Calmette, Lille / RAMES,C., Hôpital Nord, Amiens / RAULT,G., Centre de Perharidy, Roscoff / REIX,P., Hôpital Femme Mère Enfant, Bron / REMUS,N., Centre Hospitalier Intercommunal de Créteil, Créteil / RENOUIL,M., Groupe Hospitalier Sud Réunion, Saint-Pierre de la Réunion / REYNAUD-GAUBERT,M., Hôpital Nord, Marseille / SERMET-GAUDELUS,I., Necker Hôpital d'Enfants Malades, Paris / STREMLER,N., Hôpital d'Enfants de la Timone, Marseille / STORNI,V., Centre Hospitalier Bretagne Atlantique, Vannes / TASSET,C., Hôpital d'Enfants, Saint-Pierre de la Réunion / URBAN,T., Centre Hospitalier Universitaire d'Angers, Angers / VIGNERON,P., Centre Hospitalier Bretagne Sud, Lorient / WALLAERT,B., Hôpital Calmette, Lille / WEISS,L., Hôpital de Hautepierre, Strasbourg / WIZLA,N., Hôpital Jeanne de Flandre, Lille.

Patients and data collection

Since 2001, all patients with cystic fibrosis (CF) in France are evaluated at least once a year by one of 47 French hospital-based CF centers according to the national CF care recommendations (https://www.has-sante.fr/portail/jcms/c_2792719/fr/mucoviscidose).[1] Patients with CF treated in 38 of the participating CF centers from January 2004 to January 2017 were enrolled in the French CF Modifier Gene Study. As of January 1, 2017, 4,798 patients with CF had been included, which corresponded to approximately 80% of all French patients with CF.[2] Longitudinal data were obtained from electronic medical records or abstracted from the patients' paper records; retrospectively before 2004 and prospectively after January 2004. We analyzed patients born after 1985 with pancreatic insufficiency (n = 3,328) since pancreatic sufficient patients are known to have milder disease. Patients born before 1985 were excluded in order to limit selection biases due to an over-representation of patients with milder disease contributing to longer survival.[3] Information on liver status was collected at each visit.[4] When no history on CFLD was available in the patient's record, he/she was considered not to be affected. Age at onset of liver disease and severe liver disease were determined as defined below.

CFLD and severe CFLD definitions

CFLD was defined according to the European best practice guidance by Debray *et al.*[5] when at least two of the following characteristics were present: 1) abnormal physical examination: hepatomegaly and/or splenomegaly; 2) abnormalities of liver function tests defined as an increase of transaminase (alanine aminotransferase and/or aspartate aminotransferase) and/or gamma-glutamyl transpeptidase levels above normal upper limits; 3) ultrasonographic (US) evidence of liver involvement (heterogeneous echogenicity, irregular margins, or nodularity), portal hypertension (splenomegaly, increased thickness of the lesser omentum, spontaneous splenorenal anastomosis, large collateral veins, or ascites), or biliary abnormalities (bile duct dilatation). Patients with cirrhosis diagnosed by US, computed tomography (CT), and/or magnetic resonance imaging (MRI), and/or

portal hypertension [splenomegaly, hypersplenism (platelets < 150,000 10^9/L and white blood cells < 3000 10^9/L), and/or spontaneous portosystemic shunts on US], and/or esophageal varices were classified as having severe CFLD. [6,7]

Supplementary tables

Table S1: Number at risk, cumulative number of events, and cumulative incidence of cystic fibrosis related liver disease (CFLD) and severe CFLD

	CFLD			Severe CFLD		
Age (years)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)
0	3287	0	0	3287	0	0
5	2953	93	3.7	3028	12	0.5
10	2269	250	9.9	2403	59	2.3
15	1571	402	16.4	1700	125	5.3
20	917	513	22.8	1027	151	6.9
25	391	592	30.5	464	171	9.1

Abbreviations: CFLD: cystic fibrosis related liver disease

Table S2: Number at risk, cumulative number of events, and cumulative incidence of CFLD and severe CFLD according to *SERPINA1* Z genotypes

	***SERPINA1* Z GG**			***SERPINA1* Z AG/AA**		
Age (years)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)
CFLD						
0	3197	0	0	90	0	0
5	2875	88	3.6	78	5	8.1
10	2213	237	9.7	56	13	16.0
15	1529	386	16.3	42	16	22.4
20	897	493	22.6	20	20	29.2
25	380	567	30.0	11	25	46.9
Severe CFLD						
0	3197	0	0	90	0	0
5	2946	11	0.5	82	1	1.4
10	2338	57	2.2	65	2	5.5
15	1651	120	5.2	49	5	7.2
20	1002	144	6.8	25	7	11.3
25	452	164	9.1	12	7	11.3

Abbreviations: CFLD: cystic fibrosis related liver disease

Table S3: Number at risk, cumulative number of events, and cumulative incidence of CFLD and severe CFLD according to *SERPINA1* S genotypes

	***SERPINA1* S AA**			***SERPINA1* S AT/TT**		
Age (years)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)	At risk (n)	Cumulative number of events (n)	Cumulative incidence (%)
CFLD						
0	2868	0	0	412	0	0
5	2575	76	3.5	373	16	4.6
10	1971	215	9.9	295	32	8.7
15	1360	349	16.5	210	48	15.1
20	787	449	23.2	129	59	19.2
25	338	519	30.9	53	68	27.0
Severe CFLD						
0	2868	0	0	412	0	0
5	2635	11	0.5	387	1	0.3
10	2084	52	2.4	313	7	2.0
15	1475	110	5.3	223	13	4.0
20	887	135	7.2	139	14	4.6
25	403	154	9.6	60	15	5.3

Abbreviations: CFLD: cystic fibrosis related liver disease

Table S4: Patient clinical characteristics according to *SERPINA1* Z and S alleles

Clinical characteristics	*SERPINA1* Z (n=3287)			*SERPINA1* S (n=3280)		
Alleles distribution, % (n)	G:G (n=3197)	G:A (n=90)	P-value	A:A (n=2868)	T:A & T:T (n=412)	P-value
Sex (Male), % (n)	51.8% (1656)	53.3% (48)	0.77	51.6% (1479)	52.2% (215)	0.82
Current age (years), mean ± SD	16.0+/-7.7	16.31+/-7.6	0.75	16.0 +/- 7.7	16.3 +/- 7.4	0.33
Year of birth, % (n)			0.60			0.50
1986-1995	35.8% (1141)	32.2% (29)		35.8% (1023)	35.4% (145)	
1996-2005	40.3% (1283)	45.6% (41)		40.0% (1145)	42.7% (175)	
2006-2015	23.9% (763)	22.2% (20)		24.2% (692)	22.0% (90)	
Age at enrolment, % (n)			0.94			0.20
Birth	35.0% (1120)	33.3% (30)		35.3% (1012)	33.0% (136)	
1-10 years	36.1% (1153)	36.7% (33)		35.5% (1018)	40.0% (165)	
10-18 years	28.9% (924)	30.0% (27)		29.2% (838)	26.9% (111)	
European origin, % (n)	89.2% (2838)	96.6% (86)	0.024	88.4% (2523)	96.6% (396)	0.001
CF diagnosis <1 year old, % (n)	73.9% (2361)	75.6% (68)	0.72	73.7% (2115)	74.5% (307)	0.74
CFTR F508del homozygotes, % (n)	49.2% (1573)	63.3% (57)	0.009	49.3% (1414)	51.7% (213)	0.13
Meconium ileus, % (n)	13.5% (433)	11.1% (10)	0.51	13.7% (393)	11.2% (46)	0.16

Data are means ±SD or % (numbers) unless otherwise indicated. *Abbreviations*: CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator.

Table S5: Association of *SERPINA1* Z and S alleles with CFLD and severe CFLD adjusted on European origin, *CFTR* variants and meconium ileus

Association of *SERPINA1* Z and S alleles with CFLD and severe CFLD adjusted on European origin, *CFTR* variants and meconium ileus	**CFLD** HR (95%CI), P-value (Adjusted P-value*)	**Severe CFLD** HR (95%CI), P-value (Adjusted P-value*)
***SERPINA1* Z** (G:A vs G:G)	1.57 (1.04-2.36), P=0.031 (P_{adj}=0.13)	1.47 (0.69–3.15), P=0.32 (P_{adj}=1.0)
***SERPINA1* S** (T:A/T:T vs A/A)	0.87 (0.67-1.11), P=0.27 (P_{adj}=1.0)	0.66 (0.39-1.12), P=0.13 (P_{adj}=0.42)

*Bonferroni adjustment for 4 tests.

Abbreviations: CF: cystic fibrosis; CFTR: cystic fibrosis transmembrane conductance regulator, CFLD: cystic fibrosis-related liver disease; HR: hazard ratio; 95% CI: 95% confidence interval.

References

1. HAS. Mucoviscidose: Protocole national de diagnostic et de soins pour une maladie rare. Guide - Affection de longue durée. 2017.
2. Vaincre-la-Mucoviscidose, Ined. Registre Français de la mucoviscidose - Bilan des données 2016. 2017.
3. Corvol H, Thompson KE, Tabary O, le Rouzic P, Guillot L. Translating the genetics of cystic fibrosis to personalized medicine. *Transl Res.* Feb 2016;168:40-49.
4. Boelle PY, Debray D, Guillot L, Clement A, Corvol H, French CFMGSI. Cystic Fibrosis Liver Disease: Outcomes and Risk Factors in a Large Cohort of French Patients. *Hepatology.* Jul 30 2018.
5. Debray D, Kelly D, Houwen R, Strandvik B, Colombo C. Best practice guidance for the diagnosis and management of cystic fibrosis-associated liver disease. *J Cyst Fibros.* Jun 2011;10 Suppl 2:S29-36.
6. Debray D, Narkewicz MR, Bodewes F, et al. Cystic Fibrosis-related Liver Disease: Research Challenges and Future Perspectives. *J Pediatr Gastroenterol Nutr.* Oct 2017;65(4):443-448.
7. Stonebraker JR, Ooi CY, Pace RG, et al. Features of Severe Liver Disease With Portal Hypertension in Patients With Cystic Fibrosis. *Clin Gastroenterol Hepatol.* Apr 5 2016.