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

Loic Guillot[1], PhD, Pierre-Yves Boëlle[2], PhD, Dominique Debray[1,3], MD, PhD, and Harriet Corvol[1,4], MD, PhD

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

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

[3] AP-HP, Hôpital Necker Enfants Malades, Pediatric Hepatology Unit, Paris F-75007, France

[4] AP-HP, Hôpital Trousseau, Pediatric Pulmonology Department, Paris F-75012, France

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 in 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 has been associated with cystic fibrosis (CF)-related liver disease (CFLD), a common manifestation in patients with CF. We estimated the incidence of CFLD according to the *SERPINA1* genotype among 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% at age 25 compared to 30% for non-carriers. Although the increase in risk was similar for patients with severe CFLD (HR = 1.5, 95% CI = 0.7–3.2, P = 0.31) it did not reach significance due to limited number of severe cases among the Z allele carriers. No significant effect was found for the S allele.

Conclusion: CF patients carrying the *SERPINA1* Z allele experience an increased risk of developing CFLD and related complications compared to non-carriers patients. Routine *SERPINA1* Z genotyping upon CF diagnosis is warranted for selecting 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 caused by pathogenic variants in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein, a chloride channel expressed in epithelial cells.[1] The disease affects several organs, including the lungs, pancreas, intestine, and liver. Cystic fibrosis-related liver disease (CFLD) is a manifestation of CF in the liver involving a wide range of hepatobiliary abnormalities. Focal biliary cirrhosis is the most clinically relevant form of CFLD, since extension of the initially focal fibrogenic process may lead to multilobular biliary cirrhosis with subsequent portal hypertension and related complications.[2,3] Currently, multilobular cirrhosis ranks as the third cause of death after respiratory failure and transplantation-related complications in patients with CF.[4]

Although CF is recognized as a monogenic disease, considerable phenotypic diversity exists among patients with the same *CFTR* pathogenic variants.[5-7] Besides environmental factors, additional genetic modifiers have 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 further found that the risk of CFLD increased with age, with a frequency up to 32% by age 25, and was associated with several risk factors such as male sex, *CFTR* F508del homozygosity, and history of meconium ileus at birth.

SERPINA1 gene has been implicated in the development and progression of CFLD.[9] *SERPINA1* encodes the alpha-1 antitrypsin (AAT) protein, which is synthesized in large quantities by the liver and then secreted into the serum. Several pathogenic variants within the *SERPINA1* gene have been identified to cause AAT deficiency that predisposes the individual to liver disease and early-onset emphysema. The most common variants are the Z and S alleles, each caused by a single nucleotide polymorphism.[10] The Z variant is the allele overwhelmingly associated with liver disease.[11] Indeed, the Z-type AAT protein folds abnormally during biogenesis in the endoplasmic reticulum of hepatocytes, and is retained intracellularly rather than efficiently secreted, resulting in low serum AAT levels.[12] The intracellular accumulation of AAT mutant Z proteins within hepatocytes can cause

liver injury, cirrhosis, and hepatocellular carcinoma by triggering a cascade of chronic hepatocellular apoptosis, regeneration, and end-stage organ injury.[13] Individuals who are heterozygous for AAT, carrying one protease inhibitor normal M allele and one mutant Z allele (PiMZ or MZ) are generally considered asymptomatic and healthy with regard to liver disease. It is commonly accepted that those who are compound heterozygotes for the S and the Z alleles of AAT (PiSZ) may develop liver disease with identical manifestations to that of PiZZ patients, whereas liver disease is absent in PiSS individuals.[13] In CF, a case-control study revealed that carriers of the *SERPINA1* Z allele were 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, preventing determination of the actual risk of developing CFLD associated with the variant. Therefore, in the present study, we estimated the cumulated incidence of CFLD according to *SERPINA1* genotype in a cohort of French CF patients of unprecedented size (n = 3,328).

MATERIAL AND METHODS

We assembled the French CF Modifier Gene Study cohort, including CF patients treated in French CF centers since 2004 as previously described (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 group, 3,328 CF patients with pancreatic insufficiency born after 1985 were available to evaluate the 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], and patients with cirrhosis, portal hypertension, and/or esophageal varices were classified as having "severe CFLD" (see also **Supplementary Information**).[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 dbSNP (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 of 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 record, allowing for "interval censoring" between birth and age of the first report when the date of onset was not precisely known (this was the case for 142 out 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 out 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.

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 to those reported for Europeans (*SERPINA1* Z (A): 1.4% vs. 2%, *SERPINA1* S (T): 6.4% versus 6%). There were no patients homozygous for *SERPINA1 Z.* Details on the number of CF patients at risk at developing CFLD and severe CFLD, as well the cumulated number of CFLD events and of severe CFLD events for

the whole cohort and according to *SERPINA* Z and S genotypes are provided in **the supplementary information (Tables S1, S2 and S3).**

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 *SERPINA1* Z allele (HR = 1.6; 95% CI = 1.1–2.4, P = 0.019), reaching 47% at age 25 in the Z allele carriers versus 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), although it did not reach significance. In this last case, there were only seven severe cases among the Z allele carriers, making the cumulated incidence curve difficult to estimate accurately. Adjusting on *CFTR* variants and meconium ileus did not change the strength of the association.

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 the non-carriers patients.

In CF, the role of *SERPINA1* in liver disease was first identified in a two-stage case-control study including CF patients from several countries worldwide.[9] Both the initial and the replication studies showed that severe CFLD was associated with the *SERPINA1* Z allele (odds ratio of 4.72 and

3.42, respectively).[9] Our cohort of patients with CF confirms this association, but demonstrates a smaller difference in risk than reported previously. For example, using the cumulative incidence at age 25, the odds ratio for the Z allele would only be 2 in our cohort. Yet, the difference in cumulative risk with age is still clinically relevant, as *SERPINA1* Z carriers have a 50% greater risk of developing CFLD compared to non-carriers. Not surprisingly, we did not observe any association with the *SERPINA1* S allele, which is recognized to be associated with reduced levels of AAT protein but no liver manifestations.[18]

Our study has limitations related to its design and the use of medical data records as a primary source of information. However, we previously described that selection bias due to differential mortality was likely small.[8] A second potential issue is the rarity of the *SERPINA1* Z allele in our cohort, which reduced the precision of estimation despite the large cohort size. Indeed, there were only 84 carriers of the *SERPINA1* Z allele (PiMZ), and no homozygous PiZZ patients in the cohort. Furthermore, only seven of these patients experienced a severe disease form, making the cumulative incidence difficult to estimate accurately.

Obtaining a more accurate prediction of the risk of CFLD in CF patients remains an important issue, especially since the treatment commonly prescribed for its prevention, ursodeoxycholic acid, seems to have little to no effect.[8,19] Given this situation, identifying biomarkers to predict the occurrence of CFLD is fundamental to improve monitoring disease progression and assessing the effects of novel therapies such as bile salt analogues, anti-fibrotics, and CFTR correctors and potentiators. *SERPINA1* Z genotyping at the time of CF diagnosis could thus help 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 focusing 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. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society.* 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.* Aug 2016;14(8):1207-1215 e1203.
4. Rowland M, Gallagher C, Gallagher CG, et al. Outcome in patients with cystic fibrosis liver disease. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society.* 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. *Am J Transl Res.* 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. Bellis G, Delhillotte C, Lemonnier L. Registre Français de la mucoviscidose - Bilan des données 2015. 2016.

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.