SUPPLEMENTARY ONLINE CONTENT

Bartlett JR, et al. Genetic modifiers of liver disease in cystic fibrosis. Submitted to *JAMA*, 2009.

Supplementary Information:

SUPPLEMENTAL - METHODS

Patients - Initial (first) study

The CFLD patients were ascertained from 22 CF Centers from 10 countries: Australia (n=8 patients), Canada (n=17 patients), Czech Republic (n=17 patients), Germany (n=3 patients), Italy (n=28 patients), Netherlands (n=1 patient), Scotland (n=2 patients), Slovakia (n=4 patients), Turkey (n=4 patients), and U.S. (n=44 patients). Of these 128 subjects, 4 were excluded due to the CFLD diagnosis not being made until after the age of 30 years. The 843 CF (control) subjects without CFLD were ascertained from two countries (Canada, 84 patients from 32 Centers, and U.S., 759 patients from 42 Centers). By age 15 years, 83.7% of CF patients who developed CFLD were already diagnosed; therefore, CF control subjects younger than 15 years of age were excluded.

Patients - Replication (second) study

The CFLD patients were ascertained from 35 CF Centers from 10 countries: Argentina (n=5 patients), Australia (n=5 patients), Canada (n=24 patients), Chile (n=1 patient), France (n=9 patients), Ireland (n=8 patients), Israel (n=7 patients), Italy (n=14 patients), United Kingdom (n=4 patients) and U.S. (n=62 patients). Of these 139 subjects, 3 were excluded due to the diagnosis of CFLD not being made until after the age of 30 years. The 1088 CF (control) subjects without CFLD were ascertained from five countries (Canada, 391 patients from 32 Centers; Czech Republic, 30 patients; Ireland, 6 patients; Italy, 71 patients; and U.S., 590 patients from 54 Centers).

Enrollment criteria

Pertinent clinical documentation was reviewed independently by two hepatologists (P.R.D., S.C.L.) to ensure inclusion and exclusion criteria were met for CF liver disease (i.e. case report forms, radiology reports, endoscopy reports, physical examinations and clinical notes containing documentation of CFLD). When there was not consensus for any patient, these two reviewers met and additional information was requested if it would clarify the diagnosis of CFLD. No patient

was excluded because of race or ethnic background, which were self-reported. A total of 47 subjects were enrolled post liver transplant (26 in initial study, 21.0%; 21 in replication study, 15.4%). For these patients, source documentation for CFLD criteria was required from medical records prior to liver transplant. Exclusion criteria for CFLD were alcohol abuse, biliary atresia, liver cancer, portal vein thrombosis, clinically significant viral hepatitis, use of parenteral nutrition, and Wilson disease.

Distribution of the age of diagnosis of CFLD

The average first documentation of portal hypertension, i.e. diagnosis of CFLD, in the population of liver disease subjects was 10.6 (± 5.4) years. Males had an earlier age of diagnosis of CFLD than females for all subjects (males=8.5 yrs, females=10.5 years; P=0.007) and self-reported Caucasians (males=9.7 yrs, females=11.5 yrs; P=0.027). The diagnosis of CFLD was first established after the age of 30 in 7 subjects (ages of 32, 33, 35, 35, 40, 43, and 47 years), which is ≥ 4 SD from the mean of the normal distribution; therefore, these patients were excluded from the genetic analyses (4 patients from the initial study, and 3 patients from the replication study as mentioned above). Of the 260 CFLD patients, data was obtained to define/determine the year of CF diagnosis in 213 of the patients.

***CFTR* genotypes**

For the initial study, the majority of the 124 patients with CFLD had 2 identified pancreatic insufficient (PI) mutations (80.7%, n=100), most of which were DF508/DF508 (56.5%, n=70). In addition, 30 subjects carried biallelic PI mutations, including: DF508/G542X (n=4), DF508/1717-1G>A (n=3), DF508/G551D (n=2), DF508/R553X (n=2), and DF508/N1303K (n=2) and other combinations of 2 PI alleles (n=17). A few subjects had either one (n=16) or two (n=3) PI alleles without an identified *CFTR* mutation. Only 5 subjects (4%) had one “pancreatic sufficient” (PS) or variable PS allele, including G85E (n=2), R347P (n=2) and A455E (n=1).

The majority of the CF subjects in the initial control population was ascertained from the GMS Lung Study patient population[1], which accounts for the high percentage of DF508

homozygotes (92.6%). In addition, 55 subjects had biallelic PI mutations, including DF508/N1303K (n=10), DF508/G542X (n=6), DF508/G551D (n=5), DF508/621+1G>T (n=3), W1282X/W1282X (n=3), DF508/DI507 (n=2), DF508/3659delC (n=2) and other combinations of 2 PI alleles (n=24). Only 7 subjects did not have an identified *CFTR* mutation on one allele.

In the replication study, *CFTR* genotypes in CFLD subjects were similar to those in the initial study. The majority of the 136 patients with CFLD had 2 PI mutations (85.4%, n=116), and many were DF508/DF508 (45.6%, n=62). In addition 54 subjects had biallelic mutations, including DF508/G542X (n=8), DF508/G551D (n=5), DF058/621+1G>T (n=4), DF508/1717-1G>A (n=3), DF508/N1303K (n=3), DF508/R553X (n=3), DF508/R560T (n=2), DF508/W1282X (n=2), and W1282Z/W1282X (n=2), plus 22 other biallelic mutations. A few subjects had either one (n=14) or two (n=2) alleles without an identified *CFTR* mutation. Only 4 subjects (3%) had PS or variable PS mutations, including G85E (n=1), R347P (n=1), A455E (n=1), or 3849+10kbC>T (n=1).

For the 1088 replication control CF subjects without liver disease, the majority had 2 PI mutations (93.5%), most of which were DF508/DF508 (62.5%, n=685). In addition to the DF508 homozygotes, 332 subjects had biallelic PI mutations, including DF508/G551D (n=54), DF508/G542X (n=52), DF508/N1303K (n=35), DF508/621+1G>T (n=23), DF508/1717-1G>A (n=14), DF508/W1282X (n=14), DF508/R553X (n=12), and DF508/R1162X (n=11); the remaining 117 subjects had genotypes that occurred in ≤ 10 patients. The replication control population had 56 subjects with either one (n=44) or two (n=12) alleles with an unidentified *CFTR* mutation. Finally, 15 subjects had PS (or variable PS) mutations: A445E (n=5), G85E (n=4), 3849+10kbC>T (n=2), R334W (n=1), R117H (n=1), R347P (n=1), and 2789+5G>A (n=1).

Data collection

Only initials and dates of birth for the CFLD subjects were available on source documents (other identifying information was blacked out). Clinical data that were provided on standard case report forms included: self-reported race/ethnicity, pancreatic exocrine status, medical history, results of physical examination, laboratory blood work values, and abdominal radiology reports. In addition,

we requested the following procedure reports if available: liver explant pathology (from liver transplant), liver biopsy, endoscopy and colonoscopy. Clinical information was double-entered into the database to ensure accuracy.

DNA extraction and genotyping

For the majority of CFLD subjects enrolled in the study, blood was collected by venipuncture and DNA was extracted from peripheral blood leukocytes from one or two tubes of blood using standard protocols.[2] The remaining tubes were used for lymphocyte isolation, then further subjected to EBV-mediated transformation to establish immortal cell lines. DNA was isolated from the lymphoblastoid cell lines using 'Epicentre MasterPure Complete DNA and RNA purification kit' (Epicentre, Madison, WI). For a few CFLD subjects, genomic DNA was provided by the CF center enrolling the patient.

Initial study – genotyping

Mutations in *SERPINA1* (Z and S allele) were tested as described[3] or by direct sequencing. *MBL2* promoter and structural variants were determined by direct sequencing or microsphere-based genotyping (Luminex-100). *MBL2* structural (null) variants (B, C and D) were combined to construct the O/O genotype. *ACE* (D or I deletion), *GSTP1* (rs947894) and *TGFB1* (rs1800469, rs1982073, and rs1800471) were genotyped by direct sequencing or by Illumina BeadArray™ technology (San Diego, CA). For the alleles that were sequenced (*SERPINA1* Z; *MBL2* B, C and D "null" alleles), PCR products were purified either using 'QIAquick PCR purification kit' (Qiagen, Valencia, CA) or ExoSAP-IT (USB, Cleveland, OH) as per manufacturer's instructions. The purified PCR products were sequenced using 'Big Dye Terminator Cycle Sequencing Kit' and run on ABI PRISM 3100, 3130 or 310 according to the manufacturer's protocol (Applied Biosystems, Foster City, CA). *MBL2* structural ("null") variants (B, C, D) were combined to construct the O/O (null/null) genotype. The low-expression promoter variant (X) was combined with the normal structural (A) sequence to construct the XA allele, since XA produces low levels of MBL2 protein (Crosdale et al., 2000).[4] The assays performed by Illumina (*SERPINA1* S; *GSTP1*: *MBL2*

promoter variant, Y and X, which are the normal and low expression variants, respectively; *TGFB1*) used an established BeadArray (Illumina) technology [5,6] or the variants were sequenced directly by our lab. Other genetic variants (*ACE* and *CFTR*) were tested using standard methods.[7-9] For ACE, genotyping of the insertion (I) or deletion (D) polymorphism of the Alu repeat sequence in intron 16 of the *ACE* gene was performed by the method previously reported.[7] Briefly PCR products were run on a 3% agarose gel (NuSieve, Cambrex Bio Science Rockland Inc., Rockland, ME) and DNA fragments of 490 bp (insertion) and 190 bp (deletion) were visualized by staining with ethidium bromide. For *SERPINA1*, *ACE*, and *MBL2*, PCR was performed in 25 µl reaction volume containing 100 ng genomic DNA, 0.4 µM each forward and reverse primers, 1X buffer (1 mM $(NH_4)_2S0_4$, 67 mM Tris-HCl (pH8.8), 0.01% Tween-20), 1-3 mM $MgCl_2$, 100-200 µM dNTP mix and 0.75-1.25 units of AmpliTaq DNA polymerase (Perkin Elmer, Foster City, CA). PCR cycling conditions comprised of initial denaturation at 94^{o}C for 5 minutes followed by 30-35 amplification cycles (94^{o}C for 30-60 s, 55-60^{o}C for 30 s, 72^{o}C for 30-45 s) and concluded with 6-10 minutes extension at 72^{o}C. PCR primer sequences table (below) included primers used for DNA amplification.

Replication study – genotyping

For the replication study, genotyping of *SERPINA1 Z* allele and *TGFB1* codon 10 polymorphisms were performed from DNA extracted from peripheral blood leukocytes or lymphoblastoid cell lines. Genotyping was performed by sequencing from PCR-amplified product. For Z allele sequencing, PCR was performed in 25 µl reaction volume containing 100 ng genomic DNA, 0.4 µM each forward and reverse primer (see PCR primer sequences table below), 1X buffer (1 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH8.8), 0.01% Tween-20, 1.0 mM $MgCl_2$, 100 µM dNTP mix, and 0.75 units AmpliTaq DNA Polymerase (Perkin Elmer). PCR conditions comprised of initial denaturation at 94^{o}C for 5 minutes followed by 35 amplification cycles 94^{o}C for 60 s, 58.5^{o}C for 30 s, 72^{o}C for 30 s and concluded with 10 minutes extension at 72^{o}C. The purified PCR products were sequenced using 'Big Dye' and run on ABI PRISM 3100 or 3130 according to the manufacturer's protocol (Applied Biosystems).

For *TGFB1* codon 10 sequencing, PCR was performed in 25 µl reaction volume containing 100 ng genomic DNA, 0.4 µM each forward and reverse primer (see PCR primer sequences table below), 1X buffer (1 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH8.8), 0.01% Tween-20, 1.0 mM $MgCl_2$, 100 µM dNTP mix, 5% DMSO and 1.25 units AmpliTaq DNA polymerase. PCR conditions comprised of initial denaturation at 94°C for 5 minutes followed by 35 amplification cycles 94°C for 30 s, 55°C for 30 s, 72°C for 45 s and concluded with 5 minutes extension at 72°C. The purified PCR products were sequenced using 'Big Dye' and run on ABI PRISM 3100 or 3130 according to the manufacturer's protocol (Applied Biosystems).

PCR primer sequences table

ACE I/D Forward	CTG GAG ACC ACT CCC ATC CTT TCT
ACE I/D Reverse	GAT GTG GCC ATC ACA TTC GTC AGA
α1AP Z Allele Forward	CGA TGC TCT TCC CTG TTC TGA
α1AP Z Allele Reverse	GAG GGG AGA CTT GGT ATT TTG TTC
α1AP S Allele Forward	TAA CAT CCA GCA CTG TAA GAA G
α1AP S Allele Reverse	GGT TCA CCC TCC TCA GCC C
TGFβ1 Promoter Forward	ATT GGG GAC AGT AAA TGT ATG GGG T
TGFβ1 Promoter Reverse	AGG ACC AGG CGG AGA AGG CT
TGFβ1 Codon 10 and 25 Forward	TGT AAA ACG ACG GCC AGT GGG ATA CTG AGA CAC CCC CG
TGFβ1 Codon 10 and 25 Reverse	CGG GTG ACC TCC TTG GCG TAG

Statistical analysis

Stratified Fisher's exact T tests

Stratified association analyses were performed using Fisher's exact tests to assess the effects of genotype separately in men and women. Given the potential concern of population stratification, we additionally performed Fisher's exact tests restricted to Caucasian subjects in the initial, replication and combined datasets. All tests of association were performed using S.A.S. (version 9.1.3, Cary, NC).

Description of covariate adjustment in logistic regression models

We performed multivariable logistic regression models with variable levels of covariate adjustment for ethnicity, gender, *CFTR* genotype and *TGFB1* codon 10 genotype. Ethnicity was treated as a class variable with five levels, defined by subjects of Caucasian, African, Hispanic, Asian or mixed descent. *CFTR* and *TGFB1* codon 10 genotypes were each scored as a three-

level class variable. *CFTR* genotype was scored for the number of DF508 mutations (0, 1 or 2) the subject carried. *TGFB1* codon 10 genotype was scored for the number of copies of the minor allele (0, 1 or 2) the subject carried.

Population attributable risk

We estimated the population attributable risk percentage (PAR%) of CFLD associated with exposure to the *SERPINA1* Z allele risk (heterozygous) genotype using a modification of Levin's formula.[10,11] The observed absence of homozygotes for the minor (Z) allele in both our CFLD cases and controls at the *SERPINA1* Z locus facilitates the application of Levin's formula, which assumes a dichotomous exposure variable. Our study, a retrospective case-control study on CFLD, does not measure disease prevalence. However, Levin and others have shown that the population attributable risk may be estimated from a retrospective case-control study, provided the disease prevalence is low, which ensures that the rate of exposure to the risk genotype is similar in the control population and the general CF population. The total incidence of CFLD among CF patients is estimated to be approximately 3-5%[12] and we have demonstrated (see **Figure 1**) that the rate of the risk genotype at the *SERPINA1* Z locus is similar in our controls to the general population (and therefore expected to be similar to the general CF population). Applying the formula of Levin and Bertell (1978)[11] to estimate the population attributable risk, we have: $PAR\% = (Pr(AG|case) - Pr(AG|control))/(1 - Pr(AG|control)) \times 100$, where AG represents the risk (heterozygous) genotype at the *a1AP* Z locus. We obtain an estimated PAR% of 6.7% when using all subjects (initial plus replication samples) and 6.6% when using all Caucasian subjects only.

SUPPLEMENTAL - RESULTS

Genetic tests for association of *TGFB1* variants in the initial and replication studies

There was some evidence (uncorrected $p < 0.05$) for an association between CFLD and *TGFB1* (SNPs -509 and Codon 10) in the initial total population, but there was no significant evidence in only Caucasian subjects in the initial population. In the replication and combined populations, there was no association of the *TGFB1* codon 10 variant with CFLD. This lack of association could not be explained by variation in CFTR genotypes (data not shown; eTables 4-5).

Age of diagnosis of CFLD

The age of diagnosis of CFLD in male subjects (8.5 years) was significantly earlier (P=0.007) than in female subjects (10.5 years). Similarly, in self-reported Caucasians, males were diagnosed significantly earlier (P=0.027) than in females (males=9.7 years, females=11.5 years).

SUPPLEMENTAL – DISCUSSION

Clinical implications of a rare allele with high penetrance

The estimated population attributable risk for CF liver disease is 6.7% for the relatively rare *SERPINA1* Z allele. A similar population attributable risk might be obtained for a common SNP with a much weaker effect, such as is typically identified in genome wide studies of complex genetic diseases. From a clinical perspective, a rare variant with large penetrance (such as the Z allele) may be more useful, as it is now relatively easy and inexpensive to screen for genetic polymorphisms. Further, given the substantial increased risk associated for carriers of the *SERPINA1* Z allele, a clinician might be more inclined to seek and use this information.

SUPPLEMENTAL REFERENCES

1. Drumm ML, Konstan MW, Schluchter MD, et al. Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med.* 2005;353(14):1443-1453.

2. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.

3. Tazelaar JP, Friedman KJ, Kline RS, Guthrie ML, Farber RA. Detection of alpha 1-antitrypsin Z and S mutations by polymerase chain reaction-mediated site-directed mutagenesis. *Clin Chem.* 1992;38(8 Pt 1):1486-1488.

4. Crosdale DJ, Ollier WE, Thomson W, et al. Mannose binding lectin (MBL) genotype distributions with relation to serum levels in UK Caucasoids. *Eur J Immunogenet.* 2000;27(3):111-117.

5. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques.* 2002;Suppl:56-1.

6. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol.* 2003;68:69-78.

7. Ashavaid TF, Shalia KK, Nair KG, Dalal JJ. ACE and AT1R gene polymorphisms and hypertension in Indian population. *J Clin Lab Anal.* 2000;14(5):230-237.

8. Yim JJ, Park GY, Lee CT, et al. Genetic susceptibility to chronic obstructive pulmonary disease in Koreans: combined analysis of polymorphic genotypes for microsomal epoxide hydrolase and glutathione S-transferase M1 and T1. *Thorax.* 2000;55(2):121-125.

9. Wilson AG, di Giovine FS, Blakemore AI, Duff GW. Single base polymorphism in the human tumour necrosis factor alpha (TNF alpha) gene detectable by NcoI restriction of PCR product. *Hum Mol Genet.* 1992;1(5):353.

10. Levin ML. The occurrence of lung cancer in man. *Acta Unio Int Contra Cancrum.* 1953;9(3):531-541.

11. Levin ML, Bertell R. RE: "simple estimation of population attributable risk from case-control studies". *Am J Epidemiol.* 1978;108(1):78-79.

253 **12**. Welsh MJ, Ramsey BW, Accurso FJ, Cutting GR. Cystic Fibrosis. In: Scriver CR, Beaudet
254 AL, Sly WS, Valle D, eds. *The metabolic and molecular bases of inherited disease. 8th ed*.
255 New York: McGraw-Hill; 2001:5121-5188.
256

SUPPLEMENTAL FIGURE LEGENDS

eFigure 1: Age of liver disease diagnosis in CFLD subjects. Age distribution of diagnosis of liver disease (portal hypertension) in all patients with CFLD. Only 7 patients with CFLD had the diagnosis of portal hypertension made after the age of 30 yrs. These patients were excluded from the study population and the final analysis, due to being outliers (≥ 4SD from the mean) of the normal distribution.

eFigure 2: Accumulation of SERPINA1 (α1AP) protein in a CFLD subject with and without the Z allele. Histological sections from two CF subjects (homozygous for DF508 CFTR) with cirrhosis and portal hypertension (x300). Top panels: α1-antiprotease (α1AP, SERPINA1) MZ heterozygote, Bottom panels: SERPINA1 MM homozygote. Histological sections using periodic acid Schiff, diastase resistant (PAS-D) stain (left top and bottom) show changes consistent with severe CFLD in both patients, including large droplet fat in hepatocytes, numerous proliferated ductules, fibrous tissue and low grade cholangitis. However, tissue from the MZ subject, shows accumulation of PAS-D positive droplets within hepatocytes adjoining the fibrotic interface (left top), which are not observed in the MM subject with CFLD (left bottom). Immunoperoxidase stain using an antibody against SERPINA1, shows positive staining in the same area as the PAS-D positive droplets in tissue from the MZ subjects (middle top), which is not observed in the MM subject (middle bottom). It is clear that the positive stain for SERPINA1 is within hepatocytes of the MZ tissue and not sinusoidal lining cells or macrophages, which show positive CD68 staining for macrophages (right top and bottom).

eTable 1. Summary of clinical lab values for CF patients with severe liver disease

	Study	Number of patients	Normal range		> 1X to ≤ 2X		> 2X	
			%	#	%	#	%	#
Aspartate transaminase (AST)	Initial	122	23.0	28	43.4	53	33.6	41
	Replication	132	16.7	22	47.7	63	35.6	47
(range of values)			(≤ 30 U/L)		(31-60 U/L)		(> 60 U/L)	
Alanine transaminase (ALT)	Initial	116	47.4	55	35.3	41	17.2	20
	Replication	133	44.4	59	37.6	50	18.0	24
(range of values)			(≤ 40 U/L)		(41-80 U/L)		(> 80 U/L)	
Gamma glutamyl transferase (GGT)	Initial	110	24.5	27	16.4	18	59.1	65
	Replication	114	19.3	22	28.1	32	52.6	60
(range of values)			(≤ 30 U/L)		(31-60 U/L)		(> 60 U/L)	
Total bilirubin (T Bili)†	Initial	106	66.0	70	18.9	20	15.1	16
	Replication	111	70.3	78	17.1	19	12.6	14
(range of values)			(≤ 1.2 mg/dl)		(1.3-2.4 mg/dl)		(> 2.4 mg/dl)	
Albumin‡	Initial	104	49.0	51	42.3	44	8.7	9
	Replication	120	56.7	68	39.2	47	4.1	5
(range of values)			(≥ 3.5 g/dl)		(2.5-3.4 g/dl)		(< 2.5 g/dl)	
International normalized ratio (INR)	Initial	88	28.4	25	51.1	45	20.5	18
	Replication	90	32.2	29	47.8	43	20.0	18
(range of values)			(< 1.2)		(1.2-1.5)		(> 1.5)	

† Total bilirubin abnormal in 40.9% of patients (9 out of 22) in Initial Study and 38.1% of patients (8 out of 21) in Replication Study, prior to liver transplant.
‡ Albumin abnormal in 61.9% of patients (13 out of 21) in Initial Study and 50.0% of patients (10 out of 20) in Replication Study, prior to liver transplant.

eTable 2. Initial study: Prevalence of polymorphic genotypes according to CF patients with (CFLD) and without (no CFLD) liver disease, including by gender

Gene	Variant	SNP rs#	Status of liver disease	Geno-type	Patients with genotype		Geno-type	Patients with genotype		Geno-type	Patients with genotype		Number of patients	P value*
					#	%		#	%		#	%		
SERPINA1	S allele (T2313A)	17580	CFLD	AA	90	88.2	AT	12	11.8	TT	0	0.0	102	0.28
			Male		65	87.8		9	12.2		0	0.0	74	0.26
			Female		25	89.3		3	10.7		0	0.0	28	0.47
			no CFLD	AA	619	92.6	AT	49	7.3	TT	1	0.1	669	
			no CFLD male		331	91.9		29	8.1		0	0.0	360	
			no CFLD female		288	93.2		20	6.5		1	0.3	309	
	Z allele (G4627A)	28929474	CFLD	GG	110	88.7	AG	14	11.3	AA	0	0.0	124	$6.5x10^{-5}$
			Male		82	93.2		6	6.8		0	0.0	88	0.04
			Female		28	77.8		8	22.2		0	0.0	36	$6.1x10^{-5}$
			no CFLD	GG	741	97.4	AG	20	2.6	AA	0	0.0	761	
			no CFLD male		403	97.6		10	2.4		0	0.0	413	
			no CFLD female		338	97.1		10	2.9		0	0.0	348	
ACE	D/I deletion (T2313A)	N/A	CFLD	DD	43	35.0	DI	54	43.9	II	26	21.1	123	0.69
			Male		30	34.5		36	41.4		21	24.1	87	0.47
			Female		13	36.1		18	50.0		5	13.9	36	0.94
			no CFLD	DD	250	37.3	DI	300	44.7	II	121	18.0	671	
			no CFLD male		144	39.9		149	41.3		68	18.8	361	
			no CFLD female		106	34.2		151	48.7		53	17.1	310	
GSTP1	(A1375G)	947894	CFLD	AA	40	41.7	AG	41	42.7	GG	15	15.6	96	0.33
			Male		31	44.9		28	40.6		10	14.5	69	0.73
			Female		9	33.3		13	48.2		5	18.5	27	0.23
			no CFLD	AA	316	43.7	AG	331	45.8	GG	76	10.5	723	
			no CFLD male		171	43.8		173	44.4		46	11.8	390	
			no CFLD female		145	43.5		158	47.5		30	9.0	333	

MBL2	O	N/A	CFLD	AA	69	59.0	AO	42	35.9	OO	6	5.1	117	0.91
			Male		47	56.6		32	38.6		4	4.8	83	0.84
			Female		22	64.7		10	29.4		2	5.9	34	0.76
			no CFLD	AA	384	57.9	AO	248	37.4	OO	31	4.7	663	
			no CFLD male		207	57.9		138	38.5		13	3.6	358	
			no CFLD female		177	58.0		110	36.1		18	5.9	305	
	XA/O	N/A	CFLD	Other	95	82.6	XA/O	14	12.2	O/O	6	5.2	115	0.63
			Male		67	81.7		11	13.4		4	4.9	82	0.71
			Female		28	84.8		3	9.1		2	6.1	33	0.85
			no CFLD	Other	567	85.5	XA/O	65	9.8	O/O	31	4.7	663	
			no CFLD male		303	84.7		42	11.7		13	3.6	358	
			no CFLD female		264	86.6		23	7.5		18	5.9	305	
TGFB1	Promoter (C-509T)	1800469	CFLD	CC	44	39.6	CT	52	46.9	TT	15	13.5	111	0.04
			Male		34	42.5		35	43.7		11	13.8	80	0.14
			Female		10	32.3		17	54.8		4	12.9	31	0.14
			no CFLD	CC	413	49.6	CT	356	42.7	TT	64	7.7	833	
			no CFLD male		225	49.4		198	43.4		33	7.2	456	
			no CFLD female		188	49.9		158	41.9		31	8.2	377	
	Codon 10 (C29T)	1800470	CFLD	TT	33	29.5	CT	54	48.2	CC	25	22.3	112	0.01
			Male		26	32.1		36	44.4		19	23.5	81	0.04
			Female		7	22.6		18	58.0		6	19.4	31	0.09
			no CFLD	TT	343	40.7	CT	390	46.4	CC	109	12.9	842	
			no CFLD male		186	40.3		216	46.9		59	12.8	461	
			no CFLD female		157	41.2		174	45.7		50	13.1	381	
	Codon 25 (G74C)	1800471	CFLD	GG	93	83.8	GC	18	16.2	CC	0	0.0	111	0.61
			Male		67	82.7		14	17.3		0	0.0	81	0.66
			Female		26	86.7		4	13.3		0	0.0	30	0.83
			no CFLD	GG	592	85.9	GC	92	13.4	CC	5	0.7	689	
			no CFLD male		318	85.3		53	14.2		2	0.5	373	
			no CFLD female		274	86.8		39	12.3		3	0.9	316	

* All P values were calculated with the use of Fisher's exact test of comparisons of the genotypes.

289
290
291
292

eTable 3. Initial study - Caucasians: Prevalence of polymorphic genotypes according to CF patients with (CFLD) and without (no CFLD) liver disease, including by gender

Gene	Variant	SNP rs#	Status of liver disease	Geno-type	Patients with genotype		Geno-type	Patients with genotype		Geno-type	Patients with genotype		Number of patients	P value*
					#	%		#	%		#	%		
SERPINA1	S allele (T2313A)	17580	CFLD	AA	81	89.0	AT	10	11.0	TT	0	0.0	91	0.31
			Male		60	89.6		7	10.4		0	0.0	67	0.48
			Female		21	87.5		3	12.5		0	0.0	24	0.28
			no CFLD	AA	605	92.5	AT	48	7.3	TT	1	0.2	654	
			no CFLD male		326	91.8		29	8.2		0	0.0	355	
			no CFLD female		279	93.3		19	6.4		1	0.3	299	
	Z allele (G4627A)	28929474	CFLD	GG	100	90.1	AG	11	9.9	AA	0	0.0	111	$6.9x10^{-4}$
			Male		76	95.0		4	5.0		0	0.0	80	0.26
			Female		24	77.4		7	22.6		0	0.0	31	$1.0x10^{-4}$
			no CFLD	GG	726	97.4	AG	19	2.6	AA	0	0.0	745	
			no CFLD male		397	97.5		10	2.5		0	0.0	407	
			no CFLD female		329	97.3		9	2.7		0	0.0	338	
ACE	D/I deletion (T2313A)	N/A	CFLD	DD	38	34.2	DI	49	44.2	II	24	21.6	111	0.61
			Male		27	33.8		34	42.4		19	23.8	80	0.48
			Female		11	35.5		15	48.4		5	16.1	31	0.97
			no CFLD	DD	243	37.0	DI	296	45.2	II	117	17.8	656	
			no CFLD male		142	39.9		147	41.3		67	18.8	356	
			no CFLD female		101	33.7		149	49.6		50	16.7	300	
GSTP1	(A1375G)	947894	CFLD	AA	35	40.7	AG	37	43.0	GG	14	16.3	86	0.24
			Male		28	45.2		25	40.3		9	14.5	62	0.69
			Female		7	29.2		12	50.0		5	20.8	24	0.08
			no CFLD	AA	309	43.8	AG	325	46.0	GG	72	10.2	706	
			no CFLD male		167	43.5		172	44.8		45	11.7	384	
			no CFLD female		142	44.1		153	47.5		27	8.4	322	

MBL2	O	N/A	CFLD	AA	61	57.6	AO	40	37.7	OO	5	4.7	106	0.95
			Male		43	56.6		30	39.5		3	3.9	76	0.97
			Female		18	60.0		10	33.3		2	6.7	30	0.90
			no CFLD	AA	379	58.5	AO	240	37.0	OO	29	4.5	648	
			no CFLD male		204	57.8		136	38.5		13	3.7	353	
			no CFLD female		175	59.3		104	35.3		16	5.4	295	
	XA/O	N/A	CFLD	Other	86	82.7	XA/O	13	12.5	O/O	5	4.8	104	0.58
			Male		62	82.7		10	13.3		3	4.0	75	0.81
			Female		24	82.8		3	10.3		2	6.9	29	0.60
			no CFLD	Other	557	85.9	XA/O	62	9.6	O/O	29	4.5	648	
			no CFLD male		299	84.7		41	11.6		13	3.7	353	
			no CFLD female		258	87.5		21	7.1		16	5.4	295	
TGFB1	Promoter (C-509T)	1800469	CFLD	CC	41	41.4	CT	47	47.5	TT	11	11.1	99	0.20
			Male		32	44.4		32	44.5		8	11.1	72	0.41
			Female		9	33.3		15	55.6		3	11.1	27	0.22
			no CFLD	CC	408	50.0	CT	344	42.2	TT	64	7.8	816	
			no CFLD male		225	50.0		192	42.7		33	7.3	450	
			no CFLD female		183	50.0		152	41.5		31	8.5	366	
	Codon 10 (C29T)	1800470	CFLD	TT	31	31.0	CT	49	49.0	CC	20	20.0	100	0.06
			Male		24	32.9		34	46.6		15	20.5	73	0.17
			Female		7	25.9		15	55.6		5	18.5	27	0.24
			no CFLD	TT	339	41.1	CT	378	45.8	CC	108	13.1	825	
			no CFLD male		186	40.9		210	46.1		59	13.0	455	
			no CFLD female		153	41.4		168	45.4		49	13.2	370	
	Codon 25 (G74C)	1800471	CFLD	GG	84	84.8	GC	15	15.2	CC	0	0.0	99	0.82
			Male		60	82.2		13	17.8		0	0.0	73	0.63
			Female		24	92.3		2	7.7		0	0.0	26	0.81
			no CFLD	GG	579	85.9	GC	90	13.4	CC	5	0.7	674	
			no CFLD male		313	85.1		53	14.4		2	0.5	368	
			no CFLD female		266	86.9		37	12.1		3	1.0	306	

* All P values were calculated with the use of Fisher's exact test of comparisons of the genotypes.

eTable 4. Replication & combined studies: Prevalence of polymorphic genotypes according to CF patients with (CFLD) and without

Gene	Study	Variant	SNP rs#	Status of liver disease	Geno-type	Patients with genotype		Geno-type	Patients with genotype		Geno-type	Patients with genotype		Number of patients	P value*
						#	%		#	%		#	%		
SERPINA1	Replication study	Z allele (G4627A)	28929474	CFLD	GG	127	93.4	AG	9	6.6	AA	0	0.0	136	0.005
				Male		75	91.5		7	8.5		0	0.0	82	0.009
				Female		52	96.3		2	3.7		0	0.0	54	0.25
				no CFLD	GG	1062	98.0	AG	22	2.0	AA	0	0.0	1084	
				no CFLD male		571	97.6		14	2.4		0	0.0	585	
				no CFLD female		491	98.4		8	1.6		0	0.0	499	
	Combined studies	Z allele (G4627A)	28929474	CFLD	GG	237	91.2	AG	23	8.8	AA	0	0.0	260	$9.3x10^{-7}$
				Male		157	92.4		13	7.6		0	0.0	170	$1.2x10^{-3}$
				Female		80	88.9		10	11.1		0	0.0	90	$1.3x10^{-4}$
				no CFLD	GG	1803	97.7	AG	42	2.3	AA	0	0.0	1845	
				no CFLD male		974	97.6		24	2.4		0	0.0	998	
				no CFLD female		829	97.9		18	2.1		0	0.0	847	
TGFB1	Replication study	Codon 10 (C29T)	1800470	CFLD	TT	51	38.1	CT	62	46.2	CC	21	15.7	134	1.00
				Male		32	39.5		37	45.7		12	14.8	81	0.99
				Female		19	35.8		25	47.2		9	17.0	53	0.98
				no CFLD	TT	290	38.3	CT	349	46.1	CC	118	15.6	757	
				no CFLD male		171	40.7		186	44.3		63	15.0	420	
				no CFLD female		119	35.3		163	48.4		55	16.3	337	
	Combined studies	Codon 10 (C29T)	1800470	CFLD	TT	84	34.1	CT	116	47.2	CC	46	18.7	246	0.10
				Male		58	35.8		73	45.1		31	19.1	162	0.19
				Female		26	31.0		43	51.1		15	17.9	84	0.37
				no CFLD	TT	633	39.6	CT	739	46.2	CC	227	14.2	1599	
				no CFLD male		357	40.5		402	45.7		122	13.8	881	
				no CFLD female		276	38.4		337	47.0		105	14.6	718	

* All P values were calculated with the use of Fisher's exact test of comparisons of the genotypes.

eTable 5. Replication & combined studies - Caucasians: Prevalence of polymorphic genotypes according to CF patients with (CFLD) and without (no CFLD) liver disease, including by gender

Gene	Study	Variant	SNP rs#	Status of liver disease	Geno-type	Patients with genotype		Geno-type	Patients with genotype		Geno-type	Patients with genotype		Number of patients	P value*
						#	%		#	%		#	%		
SERPINA1	Replication study	Z allele (G4627A)	28929474	CFLD	GG	109	92.4	AG	9	7.6	AA	0	0.0	118	$2.2x10^{-3}$
				Male		68	90.7		7	9.3		0	0.0	75	$6.5x10^{-3}$
				Female		41	95.3		2	4.7		0	0.0	43	0.19
				no CFLD	GG	1040	97.9	AG	22	2.1	AA	0	0.0	1062	
				no CFLD male		558	97.6		14	2.4		0	0.0	572	
				no CFLD female		482	98.4		8	1.6		0	0.0	490	
	Combined studies	Z allele (G4627A)	28929474	CFLD	GG	209	91.3	AG	20	8.7	AA	0	0.0	229	$4.2x10^{-6}$
				Male		144	92.9		11	7.1		0	0.0	155	$4.8x10^{-3}$
				Female		65	87.8		9	12.2		0	0.0	74	$1.0x10^{-4}$
				no CFLD	GG	1766	97.7	AG	41	2.3	AA	0	0.0	1807	
				no CFLD male		955	97.5		24	2.5		0	0.0	979	
				no CFLD female		811	97.9		17	2.1		0	0.0	828	
TGFB1	Replication study	Codon 10 (C29T)	1800470	CFLD	TT	46	39.7	CT	52	44.8	CC	18	15.5	116	0.96
				Male		30	40.5		33	44.6		11	14.9	74	1.00
				Female		16	38.1		19	45.2		7	16.7	42	0.98
				no CFLD	TT	285	38.7	CT	338	45.9	CC	113	15.4	736	
				no CFLD male		167	41.0		181	44.5		59	14.5	407	
				no CFLD female		118	35.9		157	47.7		54	16.4	329	
	Combined studies	Codon 10 (C29T)	1800470	CFLD	TT	77	35.6	CT	101	46.8	CC	38	17.6	216	0.28
				Male		54	36.7		67	45.6		26	17.7	147	0.36
				Female		23	33.3		34	49.3		12	17.4	69	0.62
				no CFLD	TT	624	40.0	CT	716	45.8	CC	221	14.2	1561	
				no CFLD male		353	41.0		391	45.3		118	13.7	862	
				no CFLD female		271	38.8		325	46.5		103	14.7	699	

* All P values were calculated with the use of Fisher's exact test of comparisons of the genotypes.

300

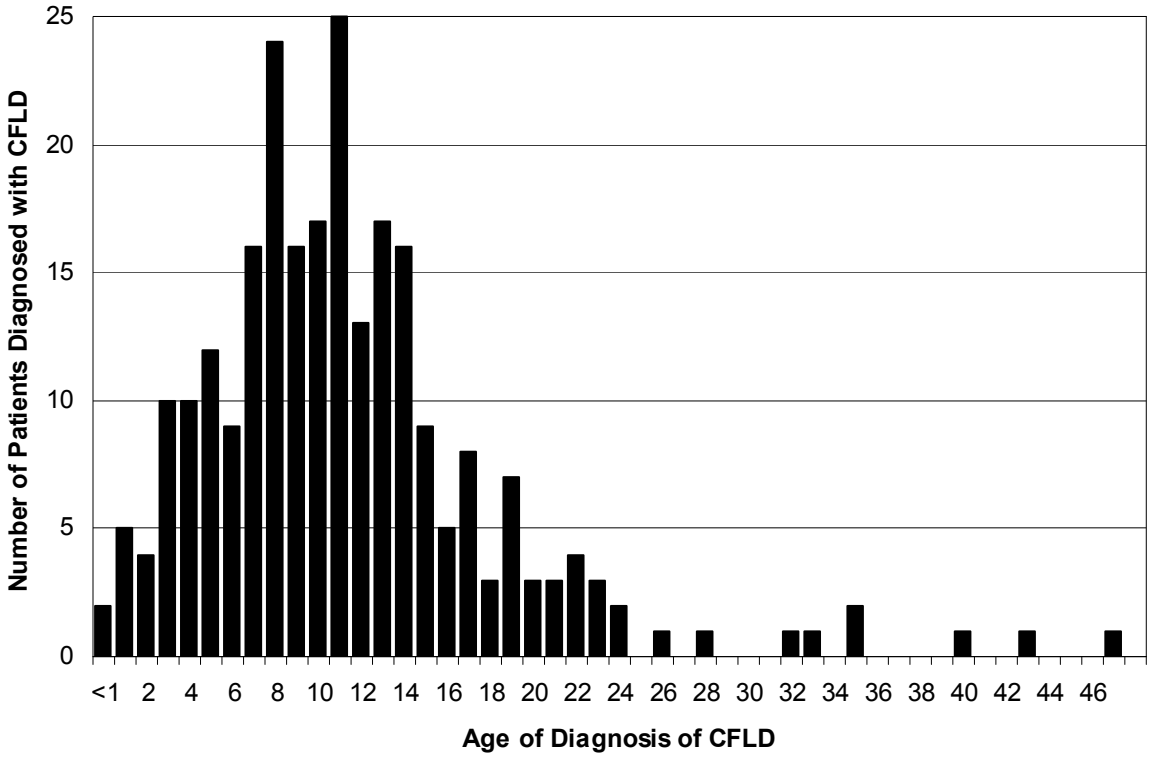
eTable 6. Multivariable logistic regression analysis of SERPINA1 Z allele among patients with CFLD and CF no LD using multiple covariates

	Odds ratio (95% CI)	P value***
Covariate - ethnicity		
All Patients (n=2105)	4.26 (2.49 - 7.28)	1.1×10^{-7}
CFLD (n = 260)		
CF no LD (n = 1845)*		
Caucasians (n=2036)	4.12 (2.37 - 7.17)	5.3×10^{-7}
CFLD (n = 229)		
CF no LD (n = 1807)		
Covariate - ethnicity, gender, CFTR and TGFB1 codon 10 genotype		
All Patients (n=1760)**	4.66 (2.55 - 8.53)	5.9×10^{-7}
CFLD (n = 246)		
CF no LD (n = 1514)		
Caucasians (n=1693)	4.53 (2.44 - 8.40)	1.8×10^{-6}
CFLD (n = 216)		
CF no LD (n = 1477)		

* CFTR genotypes unknown on 86 of the 1931 CF no LD control subjects.
** CFTR and/or TGFB1 genotypes unknown on 431 subjects of 2191 complete population.
*** All P values were calculated with the use of Fisher's exact test of comparisons of the genotypes.

301

eFigure 1. Age of liver disease diagnosis in CFLD subjects



eFigure 2. Accumulation of SERPINA1 (α1AP) protein in a CFLD subject with and without the Z allele