The reversion variant (p.Arg90Leu) at the evolutionarily adaptive p.Arg90 site in CELA3B predisposes to chronic pancreatitis

Emmanuelle Masson,1,2 Vinciane Rebourgs,3 Louis Buscaill,4 Frédérique Frete,5 Mael Pagenault,6 Alain Lachaux,7 Jean-Baptiste Chevaux,8 Emmanuelle Génin,1,2 David N. Cooper9 Claude Férec,1,2
Jian-Min Chen1,*

1Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France
2Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, F-29200 Brest, France
3Department of Gastroenterology and Pancreatology, Beaujon Hospital, Assistance Publique-Hôpitaux de Paris, Clichy, Université de Paris, Paris, France
4Department of Gastroenterology and Pancreatology, CHU Toulouse, Toulouse, France
5Service de Diabétologie – Endocrinologie, CH Libourne, Libourne, France
6Service des Maladies de l’Appareil Digestif, CHU Rennes, Rennes, France
7Service d’Hépatologie, Gastroentérologie et Nutrition pédiatriques, Hôpital Femme-Mère-Enfant, Hospices Civils de Lyon, Bron, France; Université Claude Bernard Lyon 1, Lyon, France
8Service d’Hépato-gastro-entérologie, CHRU Nancy, Vandoeuvre-les-Nancy, France
9Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, CF14 4XN, UK

Correspondence
Jian-Min Chen (jian-min.chen@univ-brest.fr) or Claude Férec (claude.ferec@univ-brest.fr)

ABSTRACT
A gain-of-function missense variant in the CELA3B gene, p.Arg90Cys (c.268C>T), has recently been reported to cause pancreatitis in an extended pedigree. Herein, we sequenced the CELA3B gene in 644 genetically unexplained French chronic pancreatitis (CP) patients (all unrelated) and 566 controls. No predicted loss-of-function variants were identified. None of the six low frequency or common missense variants detected showed significant association with CP. Nor did the aggregate rare/very rare missense variants (n=14) show any significant association with CP. However, p.Arg90Leu (c.269G>T), which was found in 4 patients but no controls and affects the same amino acid as p.Arg90Cys, serves to revert p.Arg90 to the human elastase ancestral allele. Since p.Arg90Leu has previously been shown to exert a similar functional effect to p.Arg90Cys, our findings not only confirm the involvement of CELA3B in the etiology of CP but also pinpoint a new evolutionarily adaptive site in the human genome.

KEYWORDS
CELA3B, chronic pancreatitis, gain-of-function mutation, gene conversion, elastases, paralogues

MAIN TEXT
Chronic pancreatitis (CP) is a complex disease that can be caused by genetic and/or environmental factors. Since the mapping and cloning of the first gene found to underlie hereditary pancreatitis (i.e., PRSS1; MIM# 276000) more than 20 years ago, multiple additional genes/loci associated with CP have been identified, either by means of candidate gene approaches or hypothesis-free (‘agnostic’) approaches.

CELA3B, encoding chymotrypsin-like elastase 3B (MIM# 618694), is one of the most recently identified CP-associated genes. Specifically, the whole-exome sequencing of a patient with CP, her affected daughter, unaffected brother and son, led to the identification of a missense variant in the CELA3B gene, p.Arg90Cys (c.268C>T), as the cause of the disease in a large kindred which had originally been reported over 50 years ago. Multiple lines of evidence, including experiments on CRISPR-Cas9-engineered mice, demonstrated that p.Arg90Cys gives rise to the translational upregulation of the mutant protein, which then leads to uncontrolled proteolysis and recurrent pancreatitis upon secretion and activation by trypsin.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Herein we report findings from the analysis of the CELA3B gene in 644 unrelated French CP patients and 566 controls. The patients comprised 73 cases with hereditary CP (HCP), 189 cases with familial CP (FCP) and 382 young cases (defined as either age of disease onset ≤20 years or diagnosis made at age ≤20 years, as previously described 14) with idiopathic CP (ICP). The classification of patients as HCP, FCP and ICP is in accordance with our previous publications 14,24. All participating patients had remained genetically unexplained after sequence analysis of the coding regions and flanking splice junctions of the PRSS1, SPINK1, CTRC, CFTR 25, CPA1 15, CEL-HYBV1 16 and TRPV6 21 genes. The entire coding and proximal intronic regions of the CELA3B gene were amplified using three primer pairs (Supplementary Table S1 for primer sequences). PCR was performed in a 10 µL mixture with the Expand™ Long Template PCR System (Sigma-Aldrich, Saint-Quentin Fallavier, France) according to the manufacturer’s protocol with 50 ng genomic DNA. The PCR program comprised an initial denaturation at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 15 s, annealing at 59°C for 30 s and extension at 68°C for 6 min, and a final extension at 68°C for 10 min. PCR products were purified by Illustra™ ExoProStar™ (Dominique Dutscher, Brumath, France) and then sequenced using the BigDye™ Terminator v1.1 Cycle Sequencing Kit (ThermoFisher Scientific, Waltham, MA). Sequencing primers are provided in Supplementary Table S2. We focused our analysis on (i) deletions or insertions that affected canonical GT-AG splice sites and/or coding sequence and (ii) single nucleotide substitutions that altered either canonical GT-AG splice sites or resulted in missense or nonsense variants. Variant nomenclature followed HGVS recommendations 26. NM_007352.4 was used as the reference mRNA sequence. The Brest University’s ethical review committee approved this study. All patients gave informed consent for genetic analysis.

We identified a total of 20 variants, which were classified into (i) low frequency or common (n = 6; Table 1) and (ii) rare or very rare (n = 14; Table 2) in accordance with their allele frequencies in the 566 controls. The classification of variants as very rare (allele frequency of <0.001), rare (allele frequency from 0.001 to <0.005), low frequency (from 0.005 to 0.05) and common (allele frequency of >0.05) followed Manolio and colleagues 27.

All 20 variants were predicted to result in either single or multiple missense variants. In other words, no predicted loss-of-function (pLoF) variants such as nonsense, canonical splice-site or frameshifting variants (in accordance with the gnomAD definition of pLoF variants 28) were found in any patient. This is consistent with two observations. First, the previously reported CP-causing p.Arg90Cys is a gain-of-function variant by virtue of its upregulatory effect on translation 29. Second, the pLi score for CELA3B in gnomAD (http://gnomad.broadinstitute.org/; as of 13 November 2020) is 0, suggesting that the gene is completely tolerant of heterozygous loss-of-function variants. In this regard, it is pertinent to mention that a CELA3B intronic variant, c.643-7G>T (rs61777963), manifests an association with alcoholic CP with a small protective effect (allele frequency: 13.8% in patients vs. 21.3% in controls; OR = 0.59, 95% CI 0.39 to 0.89; P = 0.01) 29. However, as acknowledged by the original authors, the number (n = 120) of alcoholic CP patients analyzed was small, and no association was found in a small cohort (n = 105) of non-alcoholic CP (allele frequency: 18.6% in patients vs. 21.3% in controls; OR = 0.84, 95% CI 0.56 to 1.26; P = 0.4) 29. We extracted corresponding data from our patients and controls, showing no significant association (allele frequency: 17.2% (222/1288) vs. 17.1% (194/1132); OR = 1.01, 95% CI 0.81 to 1.24; P = 1.0). Therefore, the aforementioned protective association is most likely spurious. Three variants, namely the common c.[71G>A;73C>T;91A>C], rare c.[529G>C;536T>G] and very rare c.736_742delACCCGCAinsTTCA(TCT, involved ≥2 closely spaced single nucleotide substitutions. The ≥2 single nucleotide substitutions in each case were confirmed to be in cis by a newly developed next-generation sequencing method (detailed method will be published elsewhere), with the original sequencing data being deposited in the NCBI Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra) under accession numbers SAMN16675587, SAMN16675586 and SAMN16675585. c.736_742delACCCGCAinsTTCA(TCT has previously been shown to be a gene conversion event 29. c.[71G>A;73C>T;91A>C] and c.[529G>C;536T>G] probably also arose via gene
<table>
<thead>
<tr>
<th>Location</th>
<th>Variant</th>
<th>HCP</th>
<th>FCP</th>
<th>ICP</th>
<th>All CP</th>
<th>Control</th>
<th>OR (95% CI); P value&lt;sup&gt;b&lt;/sup&gt;</th>
<th>rs number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 2</td>
<td>c.[71G&gt;A;73C&gt;T;91A&gt;C]&lt;sup&gt;c&lt;/sup&gt;</td>
<td>p.[Arg24His;Pro25Ser;Asn31His]</td>
<td>4/73</td>
<td>1/189</td>
<td>3/382</td>
<td>8/644</td>
<td>8/566</td>
<td>0.88 (0.33–2.35); 0.79</td>
</tr>
<tr>
<td>Exon 4</td>
<td>c.235T&gt;C</td>
<td>p.Trp79Arg</td>
<td>1/73</td>
<td>11/189</td>
<td>12/382</td>
<td>24/644</td>
<td>12/566</td>
<td>1.79 (0.89–3.61); 0.10</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.415G&gt;A</td>
<td>p.Val139Ile</td>
<td>0/73</td>
<td>2/189</td>
<td>2/382</td>
<td>4/644</td>
<td>7/566</td>
<td>0.50 (0.15–1.71); 0.26</td>
</tr>
<tr>
<td>Exon 6</td>
<td>c.625A&gt;G</td>
<td>p.Ile209Val</td>
<td>6/73</td>
<td>10/189</td>
<td>18(1)/382</td>
<td>34(1)/644</td>
<td>26/566</td>
<td>1.19(0.71–2.01); 0.50</td>
</tr>
<tr>
<td>Exon 6</td>
<td>c.629G&gt;A</td>
<td>p.Arg210His</td>
<td>4/73</td>
<td>6/189</td>
<td>20/382</td>
<td>30/644</td>
<td>21/566</td>
<td>1.27 (0.72–2.24); 0.41</td>
</tr>
<tr>
<td>Exon 7</td>
<td>c.722C&gt;G</td>
<td>p.Ala241Gly</td>
<td>3/73</td>
<td>5/189</td>
<td>15/382</td>
<td>23/644</td>
<td>21/566</td>
<td>0.96 (0.53–1.76); 0.89</td>
</tr>
</tbody>
</table>

<sup>a</sup>Number of homozygotes is indicated in parentheses wherever applicable.

<sup>b</sup>Calculation was based on allele frequency in patients vs. that in controls.

<sup>c</sup>A gene conversion event that can be alternatively termed c.71_91conNM_005747.5:c.71_91; deposited in the NCBI Sequence Read Archive database under the accession number SAMN16675587.

<sup>d</sup>The three component variants of c.[71G>A;73C>T;91A>C] have previously been registered as independent single nucleotide substitutions.

<sup>e</sup>In hg19, the reference sequence at this position is the minor C allele.

Abbreviations: CI, confidence interval; CP, chronic pancreatitis; HCP, hereditary chronic pancreatitis; FCP, familial chronic pancreatitis; ICP, idiopathic chronic pancreatitis. OR, odds ratio.
Table 2. Rare/very rare CELA3B variants in French CP patients and controls

<table>
<thead>
<tr>
<th>Location</th>
<th>Variant</th>
<th>HCP</th>
<th>FCP</th>
<th>ICP</th>
<th>All CP</th>
<th>Control</th>
<th>rs number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 3</td>
<td>c.145G&gt;A</td>
<td>p.Glu49Lys</td>
<td>1/73</td>
<td>0/189</td>
<td>0/382</td>
<td>1/644</td>
<td>0/566</td>
</tr>
<tr>
<td>Exon 4</td>
<td>c.269G&gt;T</td>
<td>p.Arg90Leu</td>
<td>0/73</td>
<td>2/189</td>
<td>2/382</td>
<td>4/644</td>
<td>0/566</td>
</tr>
<tr>
<td>Exon 4</td>
<td>c.323T&gt;A</td>
<td>p.Phe108Tyr</td>
<td>0/73</td>
<td>0/189</td>
<td>1/382</td>
<td>1/644</td>
<td>0/566</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.391C&gt;T</td>
<td>p.Arg131Cys</td>
<td>1/73</td>
<td>1/189</td>
<td>1/382</td>
<td>3/644</td>
<td>2/566</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.401A&gt;T</td>
<td>p.Gln134Leu</td>
<td>0/73</td>
<td>0/189</td>
<td>4/382</td>
<td>4/644</td>
<td>1/566</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.460G&gt;A</td>
<td>p.Glu154Lys</td>
<td>0/73</td>
<td>1/189</td>
<td>0/382</td>
<td>1/644</td>
<td>1/566</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.488G&gt;A</td>
<td>p.Gly163Asp</td>
<td>0/73</td>
<td>0/189</td>
<td>0/382</td>
<td>0/644</td>
<td>1/566</td>
</tr>
<tr>
<td>Exon 5</td>
<td>c.491G&gt;A</td>
<td>p.Arg164His</td>
<td>0/73</td>
<td>0/189</td>
<td>0/382</td>
<td>0/644</td>
<td>1/566</td>
</tr>
<tr>
<td>Exon 6</td>
<td>c.[529G&gt;C;536T&gt;G]</td>
<td>p.[Glu177Gln;Leu179Arg]</td>
<td>0/73</td>
<td>0/189</td>
<td>0/382</td>
<td>0/644</td>
<td>2/566</td>
</tr>
<tr>
<td>Exon 6</td>
<td>c.627C&gt;G</td>
<td>p.Ile209Met</td>
<td>0/73</td>
<td>0/189</td>
<td>0/382</td>
<td>0/644</td>
<td>1/566</td>
</tr>
<tr>
<td>Exon 7</td>
<td>c.682G&gt;A</td>
<td>p.Gly228Ser</td>
<td>1/73</td>
<td>0/189</td>
<td>0/382</td>
<td>1/644</td>
<td>0/566</td>
</tr>
<tr>
<td>Exon 7</td>
<td>c.736_742delACCCGC AinsTTCATCT</td>
<td>p.Thr246_Arg248delinsPheIleTrp</td>
<td>0/73</td>
<td>1/189</td>
<td>1/382</td>
<td>2/644</td>
<td>0/566</td>
</tr>
<tr>
<td>Exon 7</td>
<td>c.764G&gt;C</td>
<td>p.Arg255Pro</td>
<td>0/73</td>
<td>0/189</td>
<td>0/382</td>
<td>0/644</td>
<td>2/566</td>
</tr>
<tr>
<td>Exon 8</td>
<td>c.799A&gt;G</td>
<td>p.Ile267Val</td>
<td>1/73</td>
<td>2/189</td>
<td>3(1)/382</td>
<td>6(1)/644</td>
<td>5/566</td>
</tr>
</tbody>
</table>

Number of homozygotes is indicated in parentheses wherever applicable.

A same patient carried c.391C>T and c.736_742delACCCGC AinsTTCATCT. Whether the two variants are in cis or in trans was not determined.

A gene conversion event that can be alternatively termed c.529_536conNM_005747.5:c.529_536; deposited in the NCBI Sequence Read Archive database under the accession number SAMN16675586.

The two component variants of c.[529G>C;536T>G] have previously been registered as independent single nucleotide substitutions.

A gene conversion event that can be alternatively termed c.736_742conNM_005747.5:c.736_742; deposited in the NCBI Sequence Read Archive database under the accession number SAMN1667558.

Abbreviations: CP, chronic pancreatitis; HCP, hereditary chronic pancreatitis; FCP, familial chronic pancreatitis; ICP, idiopathic chronic pancreatitis.
conversion as, in each case, a putative donor sequence is present at the aligned positions of the highly homologous and tandemly linked CELA3A gene on human chromosome 1p36.12. It should be noted that gene conversion events involving ≥2 nucleotide substitutions are a subtype of simultaneously generated multiple nucleotide variants.

The carrier frequencies of each of the six low frequency or common missense variants are broadly similar between the HCP, FCP and ICP patients (Table 1). We therefore combined the three clinical datasets for the purposes of analysis at the individual variant level. None of the variants were found to be associated with CP in terms of a significantly different allele frequency between patients and controls. As for the rare or very rare variants (Table 2), we combined the three clinical datasets in order to perform an aggregate association analysis. 22 (3.4%) of the 644 patients and 16 (2.8%) of the 566 controls harbored rare/very rare variants, a difference which was not significant (OR = 1.22, 95% CI 0.63 to 2.34; P = 0.56).

The above notwithstanding, p.Arg90Leu (c.269G>T; Figure 1A), which affected the same amino acid as the CP-causing p.Arg90Cys, was found in 4 patients (two FCP and two ICP) but in none of the controls (Table 2). p.Arg90Leu is also absent from the 574 French subjects in the public dataset of the French Exome (FrEx) project and is extremely rare in gnomAD (allele frequency 0.0008097 in all populations). Most importantly, this variant has been previously subjected to functional characterization together with the disease-causing CELA3B p.Arg90Cys variant; these variants were remarkably similar in terms of all their measured biochemical and functional parameters as well as mouse phenotypes. It should be noted that the p.Arg90Leu variant had not been found in any patient in the original Moore study; it was functionally analyzed because, of the six human elastases, only CELA3B has an arginine at position 90 whereas all the others have a leucine. In this regard, we constructed the phylogenetic tree of the human elastase paralogues by means of NGPhylogeny.fr, thereby formally confirming that p.Leu90 represents the ancestral allele whereas p.Arg90 is the derived allele (Figure 1B). Interestingly, replacement of p.Leu90 of the human wild-type CELA3A by arginine was found to reduce protein expression. The constellation of these genetic, functional and evolutionary data therefore argues that p.Arg90 in CELA3B was an evolutionarily adaptive change and that reversion to the ancestral allele predisposes to CP.

![Figure 1](image-url)
In summary, on the basis of sequencing a large French cohort of CP patients and controls, we provide new evidence to support the involvement of the CELA3B gene in the etiology of CP. Moreover, our identification of the p.Arg90Leu in multiple CP patients has revealed a new instance in which genetic studies have helped to pinpoint evolutionarily adaptive sites. Larger genetic and functional studies are however required to determine whether other variants of CELA3B that occurred beyond the p.Arg90 site might also confer a risk for CP.

**Financial support:** This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), France. The funding source did not play any roles in the study design, collection, analysis, and interpretation of the data and in the writing of the report.

**Potential competing interests:** None.

**Data availability**

All data relevant to the study are included in the article or uploaded as supplementary information.

**REFERENCES**


