Open access • Posted Content • DOI:10.1101/2020.11.28.20240135

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

Emmanuelle Masson, Vinciane Rebours, Louis Buscail, Frédérique Frete ...+7 more authors

Institutions: French Institute of Health and Medical Research, University of Paris, University of Toulouse, Claude Bernard University Lyon 1 ...+1 more institutions

Published on: 30 Nov 2020 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Missense mutation

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

Emmanuelle Masson[1,2], Vinciane Rebours[3], Louis Buscail[4], Frédérique Frete[5], Mael Pagenault[6], Alain Lachaux[7], Jean-Baptiste Chevaux[8], Emmanuelle Génin[1,2], David N. Cooper[9], Claude Férec[1,2,*], Jian-Min Chen[1,*]

[1]Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France
[2]Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, F-29200 Brest, France
[3]Department of Gastroenterology and Pancreatology, Beaujon Hospital, Assistance Publique-Hôpitaux de Paris, Clichy, Université de Paris, Paris, France
[4]Department of Gastroenterology and Pancreatology, CHU Toulouse, Toulouse, France
[5]Service de Diabétologie – Endocrinologie, CH Libourne, Libourne, France
[6]Service des Maladies de l'Appareil Digestif, CHU Rennes, Rennes, France
[7]Service d'Hépatologie, Gastroentérologie et Nutrition pédiatriques, Hôpital Femme-Mère-Enfant, Hospices Civils de Lyon, Bron, France; Université Claude Bernard Lyon 1, Lyon, France
[8]Service d'Hépato-gastro-entérologie, CHRU Nancy, Vandœuvre-les-Nancy, France
[9]Institute 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 [1-3]. Since the mapping and cloning of the first gene found to underlie hereditary pancreatitis (i.e., *PRSS1*; MIM# 276000) more than 20 years ago [4-7], multiple additional genes/loci associated with CP have been identified, either by means of candidate gene approaches [8-17] or hypothesis-free ('agnostic') approaches [18-22].

CELA3B, encoding chymotrypsin-like elastase 3B (MIM# 618694), is one of the most recently identified CP-associated genes [20]. 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 [20] in a large kindred which had originally been reported over 50 years ago [23]. 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 [20].

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-HYB1* [16] and *TRPV6* [21] genes. The entire coding and proximal intronic regions of the *CELA3B* gene were amplified using three primer pairs (see 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 [20]. Second, the pLI score for *CELA3B* in genomAD (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_742delACCCGCAinsTTCATCT, 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 SAMN1667558. c.736_742delACCCGCAinsTTCATCT 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 1. Low frequency and common *CELA3B* variants in French CP patients and controls

Location	Variant		HCP	FCP	ICP	All CP	Control	OR (95% CI); *P* value[b]	rs number
	Nucleotide change	Amino acid change	+/n	+/n	+[a]/n	+[a]/n	+/n		
Exon 2	c.[71G>A;73C>T; 91A>C][c]	p.[Arg24His;Pro25Ser; Asn31His]	4/73	1/189	3/382	8/644	8/566	0.88 (0.33–2.35); 0.79	rs141038744; rs769262423; rs138865928[d]
Exon 4	c.235T>C[e]	p.Trp79Arg	1/73	11/189	12/382	24/644	12/566	1.79 (0.89–3.61); 0.10	rs7528405
Exon 5	c.415G>A	p.Val139Ile	0/73	2/189	2/382	4/644	7/566	0.50 (0.15–1.71); 0.26	rs141568613
Exon 6	c.625A>G	p.Ile209Val	6/73	10/189	18(1)/382	34(1)/644	26/566	1.19(0.71–2.01); 0.50	rs114365157
Exon 6	c.629G>A	p.Arg210His	4/73	6/189	20/382	30/644	21/566	1.27 (0.72–2.24); 0.41	rs112944567
Exon 7	c.722C>G	p.Ala241Gly	3/73	5/189	15/382	23/644	21/566	0.96 (0.53–1.76); 0.89	rs114895362

[a]Number of homozygotes is indicated in parentheses wherever applicable.

[b]Calculation was based on allele frequency in patients vs. that in controls.

[c]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.

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

[e]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

Location	Variant		HCP	FCP	ICP	All CP	Control	rs number
	Nucleotide change	Amino acid change	+/n	+/n	+[a]/n	+[a]/n	+/n	
Exon 3	c.145G>A	p.Glu49Lys	1/73	0/189	0/382	1/644	0/566	rs1298245114
Exon 4	c.269G>T	p.Arg90Leu	0/73	2/189	2/382	4/644	0/566	rs149443835
Exon 4	c.323T>A	p.Phe108Tyr	0/73	0/189	1/382	1/644	0/566	Pending (newly described in this study)
Exon 5	c.391C>T	p.Arg131Cys	1/73	1[b]/189	1/382	3/644	2/566	rs149805485
Exon 5	c.401A>T	p.Gln134Leu	0/73	0/189	4/382	4/644	1/566	rs4272592
Exon 5	c.460G>A	p.Glu154Lys	0/73	1/189	0/382	1/644	1/566	rs112909663
Exon 5	c.488G>A	p.Gly163Asp	0/73	0/189	0/382	0/644	1/566	rs1158940493
Exon 5	c.491G>A	p.Arg164His	0/73	0/189	0/382	0/644	1/566	rs562385324
Exon 6	c.[529G>C;536T>G][c]	p.[Glu177Gln;Leu179Arg]	0/73	0/189	0/382	0/644	2/566	rs139222231; rs148974518[d]
Exon 6	c.627C>G	p.Ile209Met	0/73	0/189	0/382	0/644	1/566	rs77941170
Exon 7	c.682G>A	p.Gly228Ser	1/73	0/189	0/382	1/644	0/566	rs141916705
Exon 7	c.736_742delACCCGCAinsTTCATCT[e]	p.Thr246_Arg248delinsPheIleTrp	0/73	1[b]/189	1/382	2/644	0/566	Previously described in Párniczky et al. (2016)[29]
Exon 7	c.764G>C	p.Arg255Pro	0/73	0/189	0/382	0/644	2/566	rs140718049
Exon 8	c.799A>G	p.Ile267Val	1/73	2/189	3(1)/382	6(1)/644	5/566	rs770756956

[a]Number of homozygotes is indicated in parentheses wherever applicable.

[b]A same patient carried c.391C>T and c.736_742delACCCGCAinsTTCATCT. Whether the two variants are in *cis* or in *trans* was not determined.

[c]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.

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

[e]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 [30] 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 [31,32].

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 [33] 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 [20]. 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 [20]. In this regard, we constructed the phylogenetic tree of the human elastase paralogues by means of NGPhylogeny.fr [34], 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 [20]. 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.

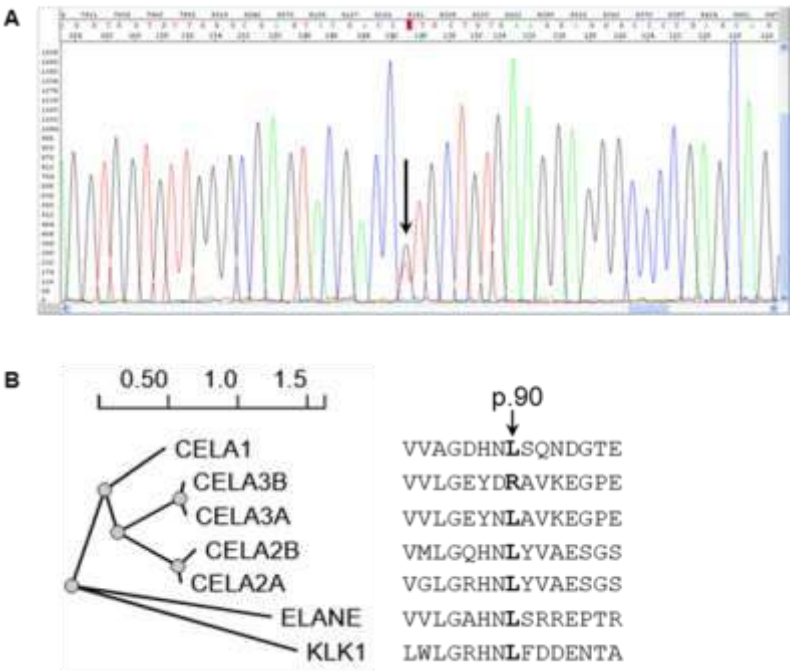

Figure 1. (**A**) Sanger sequencing electropherogram showing the CELA3B c.269G>T (p.Arg90Leu) variant (indicated by arrow) in a patient. (B) Phylogenetic tree of the human elastases. KLK1 (kallikrein 1) was used as an outgroup. Aligned amino acid sequences spanning p.90 are also shown.

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 [35,36]. 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

1. Kleeff, J. *et al.* Chronic pancreatitis. *Nat Rev Dis Primers* **3**, 17060 (2017).
2. Beyer, G., Habtezion, A., Werner, J., Lerch, M.M. & Mayerle, J. Chronic pancreatitis. *Lancet* **396**, 499-512 (2020).
3. Hegyi, P. *et al.* International consensus guidelines for risk factors in chronic pancreatitis. Recommendations from the working group for the international consensus guidelines for chronic pancreatitis in collaboration with the International Association of Pancreatology, the American Pancreatic Association, the Japan Pancreas Society, and European Pancreatic Club. *Pancreatology* **20**, 579-585 (2020).
4. Le Bodic, L. *et al.* The hereditary pancreatitis gene maps to long arm of chromosome 7. *Hum Mol Genet* **5**, 549-54 (1996).
5. Whitcomb, D.C. *et al.* A gene for hereditary pancreatitis maps to chromosome 7q35. *Gastroenterology* **110**, 1975-80 (1996).
6. Pandya, A. *et al.* Linkage studies in a large kindred with hereditary pancreatitis confirms mapping of the gene to a 16-cM region on 7q. *Genomics* **38**, 227-30 (1996).
7. Whitcomb, D.C. *et al.* Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* **14**, 141-5 (1996).
8. Cohn, J.A. *et al.* Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. *N Engl J Med* **339**, 653-8 (1998).
9. Sharer, N. *et al.* Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. *N Engl J Med* **339**, 645-52 (1998).
10. Witt, H. *et al.* Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet* **25**, 213-6 (2000).
11. Le Maréchal, C. *et al.* Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* **38**, 1372-4 (2006).
12. Witt, H. *et al.* A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. *Nat Genet* **38**, 668-73 (2006).
13. Rosendahl, J. *et al.* Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. *Nat Genet* **40**, 78-82 (2008).
14. Masson, E., Chen, J.M., Scotet, V., Le Maréchal, C. & Férec, C. Association of rare chymotrypsinogen C (*CTRC*) gene variations in patients with idiopathic chronic pancreatitis. *Hum Genet* **123**, 83-91 (2008).
15. Witt, H. *et al.* Variants in CPA1 are strongly associated with early onset chronic pancreatitis. *Nat Genet* **45**, 1216-20 (2013).
16. Fjeld, K. *et al.* A recombined allele of the lipase gene *CEL* and its pseudogene *CELP* confers susceptibility to chronic pancreatitis. *Nat Genet* **47**, 518-522 (2015).

17. Lasher, D. *et al.* Protease-sensitive pancreatic lipase variants are associated with early onset chronic pancreatitis. *Am J Gastroenterol* **114**, 974-983 (2019).
18. Whitcomb, D.C. *et al.* Common genetic variants in the *CLDN2* and *PRSS1-PRSS2* loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet* **44**, 1349-54 (2012).
19. Rosendahl, J. *et al.* Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut* **67**, 1855-1863 (2018).
20. Moore, P.C. *et al.* Elastase 3B mutation links to familial pancreatitis with diabetes and pancreatic adenocarcinoma. *J Clin Invest* **129**, 4676-4681 (2019).
21. Masamune, A. *et al.* Variants that affect function of calcium channel TRPV6 are associated with early-onset chronic pancreatitis. *Gastroenterology* **158**, 1626-1641 e8 (2020).
22. Zou, W.B. *et al. TRPV6* variants confer susceptibility to chronic pancreatitis in the Chinese population. *Hum Mutat* **41**, 1351-1357 (2020).
23. Davidson, P., Costanza, D., Swieconek, J.A. & Harris, J.B. Hereditary pancreatitis. A kindred without gross aminoaciduria. *Ann Intern Med* **68**, 88-96 (1968).
24. Chen, J.M. & Férec, C. Chronic pancreatitis: genetics and pathogenesis. *Annu Rev Genomics Hum Genet* **10**, 63-87 (2009).
25. Masson, E., Chen, J.M., Audrézet, M.P., Cooper, D.N. & Férec, C. A conservative assessment of the major genetic causes of idiopathic chronic pancreatitis: data from a comprehensive analysis of *PRSS1*, *SPINK1*, *CTRC* and *CFTR* genes in 253 young French patients. *PLoS One* **8**, e73522 (2013).
26. den Dunnen, J.T. *et al.* HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat* **37**, 564-9 (2016).
27. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
28. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
29. Parniczky, A. *et al.* Genetic analysis of human chymotrypsin-like elastases 3A and 3B (CELA3A and CELA3B) to assess the role of complex formation between proelastases and procarboxypeptidases in chronic pancreatitis. *Int J Mol Sci* **17**, 2148 (2016).
30. Chen, J.M., Cooper, D.N., Chuzhanova, N., Férec, C. & Patrinos, G.P. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* **8**, 762-75 (2007).
31. Chen, J.M., Férec, C. & Cooper, D.N. Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. *Hum Mutat* **30**, 1435-48 (2009).
32. Chen, J.M., Férec, C. & Cooper, D.N. Patterns and mutational signatures of tandem base substitutions causing human inherited disease. *Hum Mutat* **34**, 1119-30 (2013).
33. Génin, E. *et al.* The french exome (FREX) project: a population-based panel of exomes to help filter out common local variants. *Genet Epidemiol* **41**, 691 (2017).
34. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* **47**, W260-W265 (2019).
35. Chen, J.M., Montier, T. & Férec, C. Molecular pathology and evolutionary and physiological implications of pancreatitis-associated cationic trypsinogen mutations. *Hum Genet* **109**, 245-52 (2001).
36. Chimpanzee, S. & Analysis, C. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87 (2005).