Human Mutation

Discovery and Functional Annotation of Novel *PRSS1* Promoter Variants in Chronic Pancreatitis

Journal:	*Human Mutation*
Manuscript ID	humu-2016-0116
Wiley - Manuscript type:	Brief Report
Date Submitted by the Author:	10-Mar-2016
Complete List of Authors:	Boulling, Arnaud; Institut National de la Santé et de la Recherche Médicale (INSERM), U613, Abrantes, Amandine; Genetique Genomique Fonctionnelle et Biotechnologies Masson, Emmanuelle; Genetique Genomique Fonctionnelle et Biotechnologies Cooper, David; Cardiff University, Institute of Medical Genetics, College of Medicine; Robaszkiewicz, Michel; Service de Gastro-Entérologie, Centre Hospitalier Régional Universitaire (CHRU) Brest, Chen, Jian-Min; INSERM U1078 and EFS-Bretagne, Ferec, Claude; Universite Brest, ; INSERM U613,
Key Words:	chronic pancreatitis, precision medicine, promoter variant, reporter gene assay

Brief Report

Discovery and Functional Annotation of Novel *PRSS1* Promoter Variants in Chronic Pancreatitis

Arnaud Boulling,[1,2,3#] Amandine Abrantes,[2,4,5#] Emmanuelle Masson,[1,4] David N. Cooper,[6] Michel Robaszkiewicz,[5] Jian-Min Chen,[1,2,3*] and Claude Férec[1,2,3,4]

[1]Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, France; [2]Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France; [3]Etablissement Français du sang (EFS) – Bretagne, Brest, France; [4]Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Régional Universitaire (CHRU) Brest, Hôpital Morvan, Brest, France; [5]Service de Gastro-Entérologie, Centre Hospitalier Régional Universitaire (CHRU) Brest, Hôpital Cavale Blanche Morvan, Brest, France; [6]Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom

[#]These authors contributed equally to the work.

*Correspondence to: Jian-Min Chen, INSERM U1078 and EFS – Bretagne, 46 rue Félix Le Dantec, 29218 Brest, France. Tel: +33-2-98449333; Fax: +33-2-98430555; E-mail: Jian-Min.Chen@univ-brest

Contract grant sponsors: The Conseil Régional de Bretagne, the Association des Pancréatites Chroniques Héréditaires, the Association de Transfusion Sanguine et de Biogénétique Gaetan Saleun, and the Institut National de la Santé et de la Recherche Médicale (INSERM), France.

ABSTRACT: In the new age of precision medicine, interpretation of the clinical relevance of disease-associated variants, particularly those rare ones, is most challenging. Herein we illustrate the fundamental importance of functional analysis in this issue, in the context of rare *PRSS1* (encoding cationic trypsinogen) promoter variants. Gain-of-function *PRSS1* missense and copy number variants are known to cause or predispose to chronic pancreatitis (CP). A *PRSS1* promoter variant, c.-30_-28delTCC, was identified in a single individual with idiopathic CP nearly 20 years ago. It has long been speculated to increase trypsinogen expression, thereby predisposing to CP. We performed functional analysis of this variant together with three new rare variants discovered from the resequencing of the *PRSS1* promoter in a French CP cohort. Contrary to our expectation, c.-30_-28delTCC resulted in reduced rather than increased gene expression, suggesting that it may act as a protective factor rather than a risk factor for CP.

KEY WORDS: chronic pancreatitis; precision medicine; promoter variant; reporter gene assay

Chronic pancreatitis (CP) is a progressive inflammation of the pancreas that gives rise to irreversible morphological changes and impairment of both exocrine and endocrine functions [Braganza et al. 2011]. The etiology of CP is complex, with contributions from several different genetic factors having been recognized over the past two decades (see [Fjeld et al. 2015] and references therein). In particular, findings from three genes, *PRSS1* (encoding cationic trypsinogen; OMIM #276000), *SPINK1* (encoding pancreatic secretory trypsin inhibitor; OMIM #167790) and *CTRC* (encoding chymotrypsin C, which specifically degrades all human trypsinogen/trypsin isoforms [Szmola and Sahin-Tóth 2007]; OMIM #601405), have firmly established the importance of a balance between activation and inactivation of trypsinogen within the pancreatic acinar cells. Specifically, gain-of-function variants, including missense and copy number variants in the *PRSS1* gene [Le Maréchal et al. 2006; Whitcomb et al. 1996] as well as loss-of-function variants in the *SPINK1* [Witt et al. 2000] and *CTRC* [Masson et al. 2008b; Rosendahl et al. 2008] genes, cause or predispose to CP.

A large number of genetic variants have so far been reported in each of the aforementioned three CP genes (see Human Gene Mutation Database (HGMD; http://www.hgmd.org)). Some types of variant are of obvious clinical relevance by virtue of their mutational type (e.g., *PRSS1* duplication and triplication copy number mutations; splice site mutations or large deletions in *SPINK1* or *CTRC*). However, the clinical relevance of other types of variant has often to be ascertained by functional annotation. In this regard, missense mutations in all the three CP genes have been subject to systematic functional characterization (*PRSS1*, [Schnúr et al. 2014]; *SPINK1,* [Boulling et al. 2007; Boulling et al. 2012; Király et al. 2007]; and *CTRC*, [Beer et al. 2013]). By contrast, in terms of regulatory variants, only those in the *SPINK1* gene, including promoter [Boulling et al. 2011; Derikx et al. 2015b; Hegyi et al. 2016] and intronic [Zou et al. 2015; Zou et al. 2016] variants, have

been studied systematically. These functional analyses not only served to improve our understanding of the underlying pathophysiological mechanisms but more importantly, in the new age of precision medicine [Aronson and Rehm 2015; Oetting et al. 2016], have helped to establish the clinical relevance or otherwise of CP-associated variants.

No regulatory variants had been functionally characterized in *PRSS1*, the first identified CP gene, until the recent identification of a *PRSS1* promoter variant (i.e., rs4726576:C>A or c.-204C>A (Fig. 1A)) that is in perfect linkage disequilibrium with the 'CP-protective' rs10273639:C>T (i.e., c.-408C>T (Fig. 1A)) variant [Boulling et al. 2015]. This finding was made during the resequencing of the promoter region of *PRSS1* in 287 French Caucasian individuals. Given that no other study has analyzed the *PRSS1* promoter in a systematic manner, we extended the resequencing to 626 French Caucasian individuals, including 242 patients with idiopathic CP and 384 healthy controls. This yielded three additional single nucleotide substitution variants, each being found only once in either patients or controls (Table 1). All three variants were confirmed by independent PCR and sequencing. The three respective carriers were not found to carry missense or splice site variants in the *PRSS1* gene. None of the three variants are listed in either *Genetic risk factors in chronic pancreatitis* (http://www.pancreasgenetics.org/index.php) or ExAC (Exome Aggregation Consortium; http://exac.broadinstitute.org/). Informed consent was obtained from each patient and the study was approved by the Ethics Committee of Brest University.

Nearly 20 years ago, a three base-pair deletion (c.-30_-28delTCC) in the *PRSS1* gene was identified in a single individual with idiopathic CP [Férec et al. 1999]. This c.-30_-28delTCC variant was the first CP-associated regulatory variant reported in the context of the aforementioned CP genes. This variant has not so far been reported in subsequent studies and is also absent from ExAC. Thus, all three newly found variants (Table 1) and the previously known c.-30_-28delTCC represent rare variants. It should also be noted that these rare

variants, together with the functional rs4726576 [Boulling et al. 2015], are clustered within the proximal promoter region of *PRSS1* (Fig. 1A). A similar distribution pattern has also been noted for variants identified in the *SPINK1* promoter region in both CP patients and controls [Boulling et al. 2011; Hegyi et al. 2016].

The c.-30_-28delTCC variant is currently registered in OMIM (Online Mendelian Inheritance in Man; http://www.omim.org/) as one of 12 selected examples of pancreatitis-associated *PRSS1* variants (as of March 10, 2016). This variant has long been speculated to increase the level of *PRSS1* transcription, thereby predisposing to CP [Férec et al. 1999]. This idea became more plausible when *PRSS1* duplication and triplication copy number mutations were identified as a novel cause of CP [Le Maréchal et al. 2006; Masson et al. 2008a]. *In vitro* functional analysis often represents the only practical means to determine the pathogenicity of patient-derived sequence variants, particularly when they are rare [MacArthur et al. 2014]. The recent establishment of a reporter gene assay system has now made possible the functional characterization of *PRSS1* promoter variants [Boulling et al. 2015]. Functional analysis of both c.-30_-28delTCC and the newly found CP-associated c.184G>A was therefore performed as previously described [Boulling et al. 2015]. For the purpose of comparison, we also included the two variants found in controls, c.173C>T and c.147>T. In brief, we introduced all four variants separately into the pGL3-(A) luciferase vector (see Fig. 1C in [Boulling et al. 2015]), transfected the corresponding constructs into rat pancreatic acinar AR42J cells treated with dexamethasone, and determined the relative reporter gene activities against the pGL3-(A) luciferase vector.

Two variants, c.184G>A and c.173C>T, showed no significant effect on reporter gene expression. However, the other two, c.-147C>T and c.-30_-28delTCC, served to reduce reporter gene expression as compared with the wild-type promoter sequence (Fig. 1B). Contrary to our expectation, c.-30_-28delTCC reduced rather than increased reporter gene

expression. In other words, c.-30_-28delTCC might reasonably be expected to act as a protective factor for CP rather than a risk factor. c.-30_-28delTCC is located within the core promoter region of the *PRSS1* gene. Within the core promoter, an optimal distance between the TATA box and transcription start site (TSS) is often required for optimal transcription of mammalian genes [Ponjavic et al. 2006]. We surmise that c.-30_-28delTCC may negatively affect *PRSS1* transcription by altering the optimal TATA–TSS distance. A precondition for this postulate is however the presence of a TATA box upstream of the variant-affected site. A search of the *PRSS1* core promoter region using the SCAN tool from the JASPAR database (http://jaspar.genereg.net/) with the TATA-Box matrix model under default conditions yielded a TATA box motif, GTATAAAGACGAGTC (underlined is the core TATA box), at positions c.-46_-32.

We previously characterized the common c-408C>T and c.204C>A variants and showed that the latter was of functional significance [Boulling et al. 2015]. In the present study, we functionally characterized three rare variants located within the proximal promoter region of the *PRSS1* gene (i.e., c.-184G>A, c.-173C>T and c.-147C>T; Fig. 1B). The availability of these five functionally characterized variants prompted us to explore a possible correlation between the variants' effect on reporter gene expression and their potential ability to disrupt or create binding sites for transcription factors known to be expressed in the exocrine pancreas. To this end, we first screened a 41 bp tract comprising 20 bp flanking each affected nucleotide on both sides, in the context of both the reference and variant alleles, for transcription factor binding sites, using the SCAN tool from the JASPAR database with all available human JASPAR matrix models (species 9606) (http://jaspar.genereg.net/) under default conditions. We then checked all transcription factors corresponding to the predicted binding sites for expression in the exocrine pancreas by means of The Human Protein Atlas database (http://www.proteinatlas.org/). However, we were unable to draw any definite

conclusions from this attempt because of (i) the degenerate nature of transcription factor binding sites and (ii) inadequate information on the spatial and temporal expression patterns of the relevant transcription factors.

In summary, we have discovered three new rare variants in the proximal promoter region of *PRSS1* through the resequencing of French CP patients and controls. All three newly identified variants were functionally characterized together with the previously described c.-30_-28delTCC. This analysis revealed that c.-30_-28delTCC resulted in reduced rather than increased gene expression, suggesting that c.-30_-28delTCC may act as a protective factor rather than a predisposing factor for CP as originally postulated [Férec et al. 1999] . As for the three newly found rare variants, only c.147C>T, which was detected in a solitary control, exhibited a significant effect on gene expression (Fig. 1B); its negative effect on gene expression suggests that this variant may also confer protection against CP. Further, all three functional variants, viz. c.204C>A [Boulling et al. 2015] and c.-147C>T and c.-30_-28delTCC, tested here (Fig. 1B), represent loss-of-function variants in terms of *PRSS1* expression. It may be that it is much more likely for a variant to disrupt a functional transcription factor binding site than it is for it to create a new transcription factor binding site that confers an opposing function. Irrespective of the underlying biological mechanisms, our findings illustrate the fundamental importance of functional analysis in allowing us to confirm or reject the clinical relevance of disease-associated rare variants.

Disclosure statement. The authors are unaware of any conflict of interest.

References

Aronson SJ, Rehm HL. 2015. Building the foundation for genomics in precision medicine. Nature 526:336-342.

Beer S, Zhou J, Szabó A, Keiles S, Chandak GR, Witt H, Sahin-Tóth M. 2013. Comprehensive functional analysis of chymotrypsin C (CTRC) variants reveals distinct loss-of-function mechanisms associated with pancreatitis risk. Gut 62:1616-1624.

Boulling A, Le Maréchal C, Trouvé P, Raguénès O, Chen JM, Férec C. 2007. Functional analysis of pancreatitis-associated missense mutations in the pancreatic secretory trypsin inhibitor (*SPINK1*) gene. Eur J Hum Genet 15:936-942.

Boulling A, Witt H, Chandak GR, Masson E, Paliwal S, Bhaskar S, Reddy DN, Cooper DN, Chen JM, Férec C. 2011. Assessing the pathological relevance of *SPINK1* promoter variants. Eur J Hum Genet 19:1066-1073.

Boulling A, Keiles S, Masson E, Chen JM, Férec C. 2012. Functional analysis of eight missense mutations in the *SPINK1* gene. Pancreas 41:329-330.

Boulling A, Sato M, Masson E, Génin E, Chen JM, Férec C. 2015. Identification of a functional *PRSS1* promoter variant in linkage disequilibrium with the chronic pancreatitis-protecting rs10273639. Gut 64:1837-1838.

Braganza JM, Lee SH, McCloy RF, McMahon MJ. 2011. Chronic pancreatitis. Lancet 377:1184-1197.

Derikx MH, Kovacs P, Scholz M, Masson E, Chen JM, Ruffert C, Lichtner P, Te Morsche RH, Cavestro GM, Férec C, Drenth JP, Witt H, et al. 2015a. Polymorphisms at *PRSS1-PRSS2* and *CLDN2-MORC4* loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. Gut 64:1426-1433.

Derikx MH, Geisz A, Kereszturi É, Sahin-Tóth M. 2015b. Functional significance of *SPINK1* promoter variants in chronic pancreatitis. Am J Physiol Gastrointest Liver Physiol 308:G779-784.

Férec C, Raguénès O, Salomon R, Roche C, Bernard JP, Guillot M, Quéré I, Faure C, Mercier B, Audrézet MP, Guillausseau PJ, Dupont C, et al. 1999. Mutations in the cationic trypsinogen gene and evidence for genetic heterogeneity in hereditary pancreatitis. J Med Genet 36:228-232.

Fjeld K, Weiss FU, Lasher D, Rosendahl J, Chen JM, Johansson BB, Kirsten H, Ruffert C, Masson E, Steine SJ, Bugert P, Cnop M, et al. 2015. A recombined allele of the lipase gene *CEL* and its pseudogene *CELP* confers susceptibility to chronic pancreatitis. Nat Genet 47:518-522.

Hegyi E, Geisz A, Sahin-Tóth M, Derikx MH, Nemeth BC, Balazs A, Hritz I, Izbeki F, Halasz A, Parniczky A, Takacs T, Kelemen D, et al. 2016. *SPINK1* Promoter Variants in Chronic Pancreatitis. Pancreas 45:148-153.

Király O, Wartmann T, Sahin-Tóth M. 2007. Missense mutations in pancreatic secretory trypsin inhibitor (SPINK1) cause intracellular retention and degradation. Gut 56:1433-1438.

Le Maréchal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, Férec C. 2006. Hereditary pancreatitis caused by triplication of the trypsinogen locus. Nat Genet 38:1372-1374.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. Nature 508:469-476.

Masson E, Le Maréchal C, Chandak GR, Lamoril J, Bezieau S, Mahurkar S, Bhaskar S, Reddy DN, Chen JM, Férec C. 2008a. Trypsinogen copy number mutations in patients with idiopathic chronic pancreatitis. Clin Gastroenterol Hepatol 6:82-88.

Masson E, Chen JM, Scotet V, Le Maréchal C, Férec C. 2008b. Association of rare chymotrypsinogen C (*CTRC)* gene variations in patients with idiopathic chronic pancreatitis. Hum Genet 123:83-91.

Oetting WS, Brenner SE, Brookes AJ, Greenblatt MS, Hart RK, Karchin R, Sunyaev SR, Taschner PE. 2016. Pathogenicity Interpretation in the Age of Precision Medicine: The 2015 Annual Scientific Meeting of the Human Genome Variation Society. Hum Mutat.

Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A. 2006. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol 7:R78.

Rosendahl J, Witt H, Szmola R, Bhatia E, Ozsvari B, Landt O, Schulz HU, Gress TM, Pfutzer R, Lohr M, Kovacs P, Bluher M, et al. 2008. Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. Nat Genet 40:78-82.

Schnúr A, Beer S, Witt H, Hegyi P, Sahin-Tóth M. 2014. Functional effects of 13 rare *PRSS1* variants presumed to cause chronic pancreatitis. Gut 63:337-343.

Szmola R, Sahin-Tóth M. 2007. Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: identity with Rinderknecht's enzyme Y. Proc Natl Acad Sci USA 104:11227-11232.

Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, Martin SP, Gates LK, Jr., Amann ST, Toskes PP, Liddle R, McGrath K, et al. 1996. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat Genet 14:141-145.

Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, Neoptolemos JP, Lerch MM, Tector M, Sandhu BS, Guda NM, Orlichenko L, et al. 2012. Common genetic variants in the *CLDN2* and *PRSS1-PRSS2* loci alter risk for alcohol-related and sporadic pancreatitis. Nat Genet 44:1349-1354.

Witt H, Luck W, Hennies HC, Classen M, Kage A, Lass U, Landt O, Becker M. 2000. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet 25:213-216.

Zou WB, Boulling A, Masson E, Cooper DN, Liao Z, Li ZS, Férec C, Chen JM. 2015. Clarifying the clinical relevance of *SPINK1* intronic variants in chronic pancreatitis. Gut.

Zou WB, Masson E, Boulling A, Cooper DN, Li ZS, Liao Z, Férec C, Chen JM. 2016. Digging deeper into the intronic sequences of the *SPINK1* gene. Gut.

Figure Legend

Figure 1. Discovery and functional characterization of *PRSS1* promoter variants. **A:** Representation of the *PRSS1* promoter region that was resequenced (i.e., extending from c.-530 to c.-1). The three newly detected rare variants (i.e., c.-184G>A, c.-173C>T and c.-147C>T) and a previously described rare variant (i.e., c.-30_-28delTCC [Férec et al. 1999]), that were functionally characterized in this study, are highlighted in bold. The 'CP-protective' rs10273639 (i.e., c.-408C>T [Derikx et al. 2015a; Whitcomb et al. 2012]) and the functional rs4726576 (i.e., c.-204C>A [Boulling et al. 2015]) variants, are shown in normal font; these two common polymorphisms were found to be in perfect linkage disequilibrium in French Caucasians [Boulling et al. 2015].. The promoter region contained within the pGL3-(A) reporter gene vector [Boulling et al. 2015] (used for constructing the variant reporter gene vectors in this study) and the proximal promoter of the *PRSS1* gene are indicated by long and short dotted lines, respectively. ×, translational initiation codon of *PRSS1*. TSS, transcription start site. **B:** Relative promoter activities of the variant constructs in relation to the PGL3-(A) construct (wild-type), in AR42J cells treated with dexamethasone. Results represent means ± SD of 3 to 6 independent experiments. **P*<0.05; ****P*<0.0001.

Table 1. Variants Detected in Resequencing the *PRSS1* Promoter Region

Variant[a]	Patients	Controls	GenBank accession number
c.-184G>A	1/242	0/384	Pending
c.-173C>T	0/242	1/384	Pending
c.-147C>T	0/242	1/384	Pending

[a]Variant nomenclature followed HGVS recommendations (http://www.hgvs.org/mutnomen/recs.html), with the A of the translational initiation codon ATG of the *PRSS1* gene being given as c.1. GenBank accession numbers NC_000007.14 and NM_002769.4 were used as reference sequences.

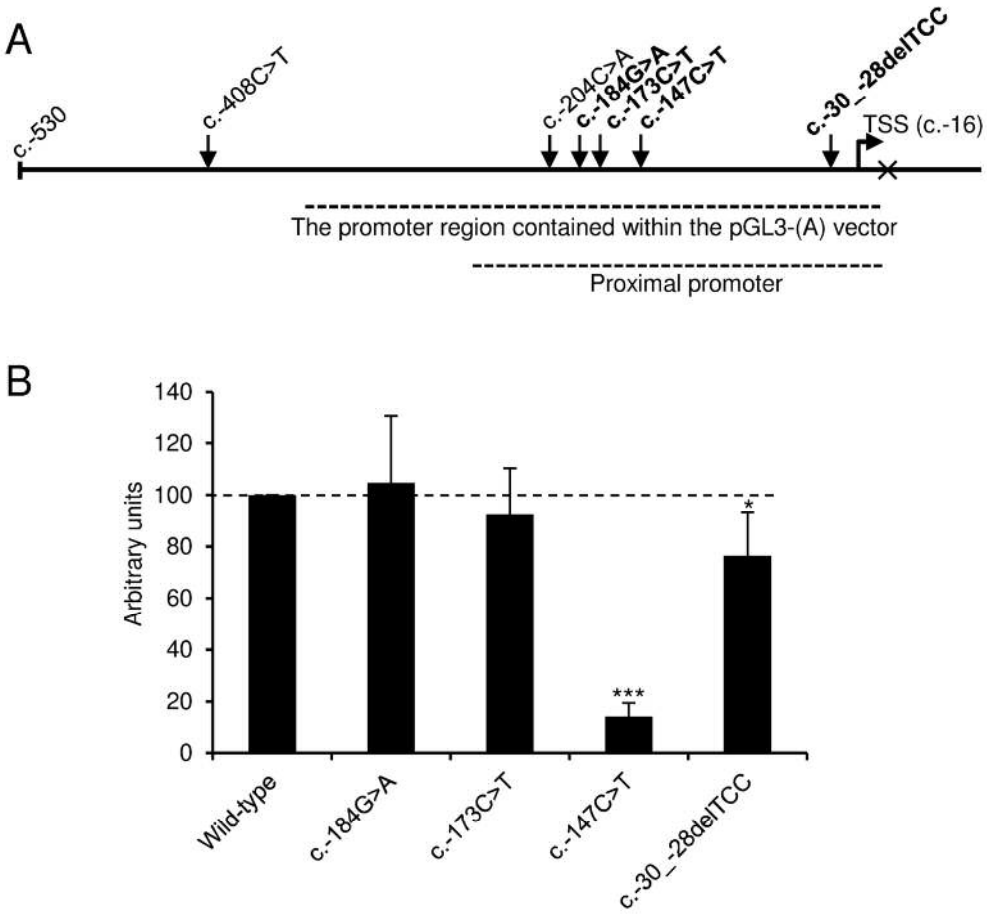

Figure 1
110x102mm (600 x 600 DPI)