Trypsinogen (*PRSS1* and *PRSS2*) Gene Dosage Correlates with Pancreatitis Risk across Genetic and Transgenic Studies: A Systematic Review and Re-Analysis

Wen-Bin Zou, Zhuan Liao

Department of Gastroenterology, Changhai Hospital, the Secondary Military Medical University, Shanghai, China

Shanghai Institute of Pancreatic Diseases, Shanghai, China

David N. Cooper

Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom

Emmanuelle Masson, Na Pu, Claude Férec, Jian-Min Chen

Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France

Service de Génétique Médicale et de Biologie de la Reproduction, CHRU Brest, F-29200 Brest, France

(Authors from Invitae pending; order of authors to be decided)

Brandie: names, affiliations and email address of authors from Invitae, conflict of interest statements, financial supports.

Correspondence

Jian-Min Chen, INSERM UMR1078 – EFS – UBO, 22 avenue Camille Desmoulins, 29238 BREST, France.

Email: jian-min.chen@univ-brest.fr

Abstract (maximum of 260 words)

BACKGROUND AND AIMS: Trypsinogen (*PRSS1* and *PRSS2*) copy number gains and regulatory variants have both been proposed to elevate pancreatitis risk through a gene dosage effect (i.e., by increasing the expression of the wild-type protein). However, to date, their impact on pancreatitis risk has not been thoroughly evaluated whilst the underlying pathogenic mechanisms remain to be explicitly investigated in mouse models.

METHODS: Genetic studies of the rare trypsinogen duplication and triplication variants, and the common rs10273639C variant, were collated from PubMed and/or ClinVar. Mouse studies that analyzed the influence of a transgenically expressed wild-type human *PRSS1* or *PRSS2* gene on the development of pancreatitis were identified from PubMed. Genetic effects of the different risk genotypes, in terms of odds ratios, were calculated wherever appropriate. Genetic effects of the rare trypsinogen duplication and triplication variants were also evaluated by reference to their associated disease subtypes.

RESULTS: We demonstrate an unambiguous positive correlation between increased trypsinogen gene dosage and pancreatitis risk in the context of the rare duplication and triplication CNVs, and between the level of trypsinogen expression and disease risk in the context of the heterozygous and homozygous rs10273639C-tagged genotypes. We retrospectively identify three mouse transgenic studies that are informative for the pathogenic mechanism underlying the trypsinogen gene dosage effect in pancreatitis.

CONCLUSION: Trypsinogen gene dosage correlates with pancreatitis risk across genetic and transgenic studies, highlighting the fundamental role of the level of wild-type trypsinogen in the etiology of pancreatitis. We propose that specifically downregulating trypsinogen expression in the pancreas could serve as a potential therapeutic and prevention strategy for pancreatitis.

Keywords: Chronic Pancreatitis; Copy Number Variant; Regulatory Variant; Gene Dosage; Genetic Effect; Transgenic Mouse Model

Main text: maximum of 4000 words.

Introduction

Chronic pancreatitis (CP) is a chronic inflammatory process that leads to progressive morphological and functional changes of both the exocrine and endocrine portions of the pancreas.[1] It has a prevalence of 36-125 per 100,000 individuals in various countries.[2] CP patients often have a poor quality of life,[3] an increased risk of pancreatic cancer,[4] and higher morbidity than the general population.[5] Since once initiated the disease is currently irreversible,[6, 7] understanding its underlying mechanisms of pathogenesis is a prerequisite for developing improved options in terms of prevention and treatment.

In 1996, a gene for autosomal dominant hereditary pancreatitis was mapped to chromosome 7q34,[8-10] followed by the identification of a causal gain-of-function missense variant, p.Arg122His, in the *PRSS1* gene (encoding cationic trypsinogen, the most abundant isoform of trypsinogen).[11] This landmark finding provided strong support for the then century-old idea that pancreatitis was an autodigestive disease due to premature trypsinogen activation within the pancreas[12] and heralded a new era in the study of CP. To date, more than 40 missense variants in the *PRSS1* gene have been reported in the literature[13] many of which have been subjected to functional analysis either in the context of biochemical characterization or transfection studies.[14-22] Several missense variants, including p.Arg122His, have also been studied in mouse models.[23-29] The pathogenic effects of a given *PRSS1* missense variant, when evident, have invariably been attributable to the "qualitative" change conferred by the variant on the mutant trypsinogen/trypsin.

Ten years after the 1996 landmark finding, a trypsinogen triplication copy number variant (CNV) involving both *PRSS1* and *PRSS2* (encoding anionic trypsinogen, the second major trypsinogen isoform) was reported as a new cause of autosomal dominant hereditary pancreatitis.[30] This study provided the first evidence that trypsinogen is subject to a dosage

effect, a finding that was further supported by the subsequent identification of CP-associated trypsinogen duplication CNVs[31] as well as a common risk-associated single nucleotide polymorphism (rs10273639C) located 408 bp upstream of the translation initiation codon of the *PRSS1* gene.[32] Quite distinct from the aforementioned "qualitative" missense variants, trypsinogen CNVs and regulatory variants have been assumed to act quantitatively by elevating CP risk through increasing the expression of the wild-type (WT) protein.[33] However, to date, the impact of trypsinogen dosage on the clinical phenotype conferred by these "quantitative" variants has not been thoroughly evaluated across different genetic platforms. More importantly, unlike "qualitative" *PRSS1* missense variants whose underlying pathogenic mechanisms have been individually and extensively studied by means of *in vitro* and *in vivo* mouse models, precisely how "quantitative" trypsinogen variants contribute to CP is still unclear. These shortcomings have hampered our understanding of the general role of trypsinogen in the etiology of CP from a mechanistic standpoint.

Herein, we have sought to address the aforementioned deficiencies through a critical assessment of currently available genetic and transgenic data. We show that (i) a general relationship between trypsinogen dosage and genetic effect is discernible from published genetic data and (ii) the results of several transgenic mouse studies are potentially informative with respect to the pathogenic mechanisms underlying the "quantitative" trypsinogen variants.

Methods

Literature Search

PRSS1-related publications listed in the Genetic Risk Factors in Chronic Pancreatitis Database (last modified on 20 December 2018)[13] were used as the starting point for screening relevant human genetic and mouse transgenic studies. This was augmented by cross-reference

checking and a keyword search (i.e., “PRSS1 duplication”, “PRSS1 triplication”, “PRSS1 copy number”, “PRSS1 CNV”, “trypsinogen duplication”, “trypsinogen triplication”, ‘trypsinogen copy number” and “trypsinogen CNV” in the context of trypsinogen copy number gain-related genetic studies and “mouse pancreatitis PRSS1” and “mouse pancreatitis transgenic trypsinogen” in the context of transgenic mouse studies) in “All Fields” of PubMed. Eligible human genetic studies refer to those describing either new trypsinogen copy number gains in CP patients or known trypsinogen copy number gains in patients with a different disease subtype (see below). It should be noted that the word “new” refers either to specific events as defined by distinct breakpoints or incompletely characterized events that have been reported in different countries. Eligible mouse transgenic studies refer to those that analyzed the effect of a transgenically expressed WT human *PRSS1* or *PRSS2* gene on the development of pancreatitis in mice by reference to control mice. The literature search was frozen on 28 September 2021.

A literature search was not however performed with respect to rs10273639C, which has recently been subject to meta-analysis.[34]

The conduct and reporting of this study were essentially in accordance with the MOOSE guidelines.[35]

Screening for Trypsinogen Copy Number Gains Deposited in ClinVar

“*PRSS1*” and “*PRSS2*” were used separately for a keyword search in ClinVar.[36] All resulting entries were manually reviewed. Copy number gains that were reported (i) to involve all five exons of *PRSS1* and (ii) to be associated with pancreatitis were extracted.

Original submitters to ClinVar (Invitae and Centogene AG) were contacted to obtain information on variant detection method, ethnicity and family history of the corresponding

carriers. Invitae re-classified their patients as having hereditary CP (HCP), familial CP (FCP) or idiopathic CP (ICP) as described below.

Disease Subtype Definitions

CP was classified into three subtypes, HCP, FCP and ICP, as previously described.[30, 37, 38] Specifically, HCP is defined in terms of having three or more affected family members spanning at least two generations whereas FCP is indicated by a positive family history but without satisfying the strict diagnostic criteria for HCP. ICP is specified when neither a positive family history of pancreatitis nor any external causative risk factors (e.g., excessive alcohol consumption, infection, trauma and drug use) have been reported. It should be noted that 'non-alcoholic CP', a term used in some publications, was regarded as being equivalent to ICP, as has been our previous practice.[39]

Public Databases Used to Evaluate the Carrier Frequencies of the Studied Variants in Normal Populations

The Genome Aggregation Database (gnomAD)[40] was used for this purpose.

Definitions of Genetic Effect

Assessment of the genetic effect was made essentially following the framework established by Manolio and colleagues; the strength of genetic effect was defined respectively as low (odds ratio (OR) < 1.1), modest (OR from 1.1 to 1.5), intermediate (OR from >1.5 to 3.0) or high (OR > 3.0).[41] To this end, ORs associated with the heterozygous or homozygous rs10273639C genotype were subjected to meta-analysis by means of the Review Manager 5.3 software,[42] as previously described.[34] OR associated with the trypsinogen duplication CNV was estimated by means of the GIGA calculator using default parameters.[43] The genetic

effects of the rare trypsinogen copy number gains were additionally evaluated by reference to their affected disease subtypes.

Results

Dosage Effect in the Context of the Rare Trypsinogen Triplication and Duplication CNVs

Six PubMed-indexed studies were identified using the criteria described in Methods (Table 1). The triplication CNV and five (numbered #1-5) of the seven duplication CNVs were identified by means of quantitative fluorescence multiplex PCR (QFM-PCR), followed by characterization of their associated breakpoints at the nucleotide sequence level.[30, 31, 33, 44] These fully characterized CNVs were all reported by our genetics laboratory in Brest. The nature of these triplication and duplication CNVs was further confirmed by fluorescence *in situ* hybridization in interphase and metaphase cells derived from one triplication carrier and one duplication carrier.[30, 31] The triplication and the #1 duplication had both arisen from a common founder chromosome.[31, 44] Despite their variable sizes (between ~400 kb and ~550 kb), the duplicated segments invariably involved the *cis*-linked *PRSS1* and *PRSS2* genes on chromosome 7q34.[33] By contrast, the two duplication CNVs reported from outside of the Brest genetics laboratory[45, 46] were not fully characterized (Table 1). The duplication CNV reported by LaRush et al.[45] first came to attention by virtue of an increase in average coverage of *PRSS1* in whole-exome resequencing and was then validated by a CNV assay specifically designed to target *PRSS1* exon 5. The other duplication CNV was detected in one of 75 Chinese children with ICP by means of a commercially available CNV assay that also targeted *PRSS1* exon 5 specifically.[46]

We searched for additional trypsinogen copy number gains in ClinVar.[36] The keyword search using "*PRSS1*" yielded a total of 311 entries. Of these, 7 represented trypsinogen copy

number gains that were described as (i) encompassing the entire coding sequence of the *PRSS1* gene and (ii) being associated with "hereditary pancreatitis". The sole example of a triplication CNV listed by ClinVar (VCV000011884.2; submitted by OMIM - Online Mendelian Inheritance in Man) actually refers to our previously published triplication CNV.[30] All of the remaining six entries (five from Invitae and one from Centogene AG; Table 2) were originally described as duplications. In all these six entries, no information on *PRSS2* was available, and a search for "*PRSS2*" in ClinVar[36] also found no *PRSS2* copy number gains in association with "hereditary pancreatitis". We contacted the original submitters with respect to clinical testing protocols; Invitae confirmed that only *PRSS1* was targeted for trypsinogen copy number detection whilst Centogene AG has not yet replied. Moreover, for each of their five entries, Invitae provided information on the number and ethnicities of the affected unrelated pancreatitis patients at the time of data submission to ClinVar. Furthermore, Invitae redefined their "hereditary pancreatitis" patients as having HCP, FCP or ICP. Finally, Invitae also identified triplication CNVs (Table 1).

Put at its simplest, the triplication and duplication CNV genotypes would be expected to generate respectively twice as much and 1.5 times as much WT trypsinogen as compared to the WT genotype (Figure 1). To explore a possible difference in genetic effect conferred by the presumed difference in increased gene dosage, we evaluated the spectrum of disease subtypes affected by the duplication and triplication CNVs. The reasoning underlying this attempt was that, the more severe the functional impact of the copy number gain, the more frequently it should be found in HCP patients. Triplication CNVs were firstly identified in five of 34 French HCP families that had not previously been found to carry any known intragenic disease-associated mutations in four known CP genes (i.e., *PRSS1*, *PRSS2*, *SPINK1* and *CFTR*). In all five triplication-positive HCP families, the triplication allele segregated perfectly with the disease (i.e., all HCP patients available for genetic analysis

carried the triplication)[30]; of the 18 triplication carriers, 16 were symptomatic, corresponding to a penetrance of 89%.[30] As shown in Table 2, a trypsinogen triplication CNV was also identified in an Hispanic family with HCP. Previous analysis of a total of 282 healthy French controls by means of QFM-PCR had not identified any trypsinogen copy number gains.[30, 31] This small number of controls was superseded by the advent of gnomAD SVs v2.1,[40, 48] with which a keyword search using "*PRSS1*" failed to detect any trypsinogen triplication CNVs in 10,846 subjects. These observations, taken together, firmly established the pathogenicity of the triplication CNV in the context of HCP.

By contrast, trypsinogen duplication CNVs were detected more rarely in HCP families (Tables 1 and 2). Most importantly, whenever they were detected in HCP,[45, 47] they were unable on their own to account for the disease. Specifically, the duplication CNV reported by LaRush et al.[45] was only found in the youngest patient, III-6, of a large HCP family. It should be noted that in this family, all five genetically tested patients (including III-6) carried *SPINK1* c.27delC, a clearly pathogenic variant that had been previously found in an HCP family as well as in an FCP family.[49] [*SPINK1* encodes pancreatic secretory trypsin inhibitor; a diverse range of loss-of-function variants in the gene have been found in CP patients.[13]] Here it is pertinent to mention that VCV000417552.1 was found in an ICP patient, who additionally carried *PRSS1* p.Arg122His (Table 2). Interestingly, this patient's mother, sister and daughter all carried the duplication (but not *PRSS1* p.Arg122His) and were not clinically affected, suggesting quite low penetrance for the duplication in this particular family.

Let us now turn to the "hybrid" trypsinogen gene or "double gain-of-function" identified in a large French HCP family affecting 6 patients across three generations.[47] The additional (and hybrid) gene copy, comprising exons 1 and 2 derived from *PRSS2* and exons 3-5 derived from *PRSS1*, segregated with the disease in this family exhibiting a penetrance of 86%. This hybrid gene copy acts concomitantly as a 'quantitative' trypsinogen duplication CNV and a

'qualitative' *PRSS1* missense variant (i.e., p.Asn29Ile). As *PRSS1* p.Asn29Ile has been reported in multiple HCP families worldwide and actually represents the second most frequently reported disease-causing variant in the *PRSS1* gene,[13, 50] the high penetrance of the disease in the French HCP family[47] is likely to be only partially attributable to the increased trypsinogen dosage conferred by the additional "hybrid" trypsinogen copy.

Finally, we attempted to quantify the genetic effect of the duplication CNV on CP risk in terms of OR values. To this end, the detection frequency of the duplication CNV in a French ICP cohort (i.e., 4/1246) was taken from Masson et al.[31] A keyword search using "*PRSS1*" in gnomAD SVs v2.1[40, 48] succeeded in identifying a single heterozygous duplication in 1 of 10,846 subjects; the duplicated segment comprised 773 kb (extending from chr7:141933998 to 142707100 (hg19)) and encompassed both the *PRSS1* and *PRSS2* loci. Employing these population data as a normal control, the OR of the duplication CNV for ICP was estimated to be 34.93 (95% confidence interval (CI): 3.90-312.74; $P = 0.00074$).

The above findings, taken together, lead to the conclusion that trypsinogen triplication CNVs cause HCP but trypsinogen duplication CNVs do not [JIAN-MIN: *PRSS1* gene duplication may well be contributory but it is neither necessary nor sufficient for clinical symptoms to occur. Is it worth pointing out that when it occurs in concert with a qualitative *PRSS1* missense variant, it can increase the genetic effect associated with the missense variant on its own, thereby making it more likely that the pancreatitis will come to clinical attention in a given invidual?], thereby making a clear distinction between them in terms of their genetic effect. The OR of the trypsinogen duplication CNV for ICP was 34.93, which translates into a "high" genetic effect [JIAN-MIAN: this sounds inconcistent with the first sentence of the paragraph] in accordance with Manolio et al.[41]. An OR for the trypsinogen triplication CNV cannot be accurately determined but it is safe to assume that it has a "very

high" genetic effect. Therefore, there is a correlation between gene dosage and genetic effect in the context of trypsinogen triplication and duplication CNVs (Figure 1).

A Dosage Effect in the Context of the Common rs10273639C Allele

In addition to the copy number gains, another type of variant that could potentially lead to the increased expression of WT trypsinogen would be regulatory variants. In this regard, the first genome-wide association study of CP led to the association of a common variant located at the *PRSS1* locus, rs10273639C, with both ICP and alcoholic CP.[32] Our current knowledge about this association may be summarized as follows. First, rare known CP-associated *PRSS1* missense variants were excluded as a possible explanation for the observed association.[32] Second, rs10273639C (located 408 bp upstream of the translation initiation codon of *PRSS1*) is in high linkage disequilibrium with rs4726576C, which is located 204 bp upstream of the *PRSS1* translation initiation codon; and a promoter reporter gene assay has shown that it is rs4726576C rather than rs10273639C that is responsible for the elevated gene expression.[51] Third, the rs10273639C-tagged risk allele is associated with slightly increased *PRSS1* and *PRSS2* mRNA expression in pancreatic tissue in a dosage-dependent manner.[32, 34] Finally, the association with ICP in three independent cohorts (i.e., German,[52] Japanese[53] and Indian[54]) was best explained by an additive genetic model in which the genetic effect of the homozygous rs10273639C genotype was approximately double that of the heterozygous rs10273639C genotype.[34]

In the present study, using genotype distribution data in ICP patients and controls from the three aforementioned studies,[52-54] we performed a meta-analysis to calculate the pooled OR of the heterozygous rs10273639C genotype together with that of the homozygous rs10273639C genotype, both relative to the homozygous rs10273639T genotype (Supplementary Figures S1 and S2). The homozygous rs10273639C genotype had an OR of

1.62, corresponding to an intermediate genetic effect, whilst the heterozygous rs10273639C genotype had an OR of 1.38, which equates to a more modest genetic effect (Figure 1).

Taken together, increased trypsinogen expression resulting from heterozygous and homozygous s10273639C-tagged genotypes correlates positively with an increased risk of CP (Figure 1).

Dosage Effect in the Context of Mice Transgenically Expressing a WT Human Trypsinogen Gene

As far as we are aware, no mouse studies have so far been performed with the explicit intention of investigating the pathogenic mechanisms underlying the "quantitative" trypsinogen variants. However, perusal of the literature identified four mouse studies that analyzed the effects of transgenically expressed WT human *PRSS1* or *PRSS2* gene,[25, 28, 29, 55] which may be retrospectively viewed as having served as *in vivo* models of trypsinogen copy number gains. One of these studies did not compare findings from the transgenic mice with those from control mice[28] and was therefore excluded from further consideration.

Of the remaining three studies, two transgenically expressed human WT *PRSS1*[25, 29] whereas the other transgenically expressed human WT *PRSS2*.[55] In humans, significant differences exist between the PRSS1 and PRSS2 proteins in terms of their expression levels in the pancreas, their biochemical properties and pathophysiological roles.[56, 57] These differences notwithstanding, the transgenically expressed WT *PRSS1* and *PRSS2* genes were regarded as being functionally equivalent here because both led to the expression of WT human *trypsinogen* in mice.

The Athwal et al. study[25] compared overall transgene expression levels between homozygous and heterozygous mice by means of Western blotting and immunohistochemistry but found no substantial differences. As opined by the original

authors, a feedback regulation process may have limited the amount of trypsinogen/trypsin expression in individual pancreatic acinar cells and/or those acinar cells that express the transgene most highly may have been subject to self-destruction.

Most importantly, in the study by Athwal et al.,[25] pathological changes characteristic of pancreatitis occurred spontaneously in up to 10% of aging (>9 month) mice harboring the transgenically expressed WT human *PRSS1* gene (termed *PRSS1*^{WT}) but not in the control mice. Further, the *PRSS1*^{WT} mice were sensitized to caerulein-induced pancreatitis. In the Huang et al. study,[29] a subtle increase of focal inflammation was observed in homozygous *PRSS1*^{WT} mice as compared to heterozygous *PRSS1*^{WT} mice, and more severe caerulein-induced pancreatitis was observed in the *PRSS1*^{WT} mice as compared to control mice. In the Wan et al. study,[55] the authors generated two lines of *PRSS2*^{WT} mice using a full-length human *PRSS2* gene-containing bacterial artificial chromosome, with line #2 expressing a ~2 fold higher level of *PRSS2*^{WT} than line #1. Focal spontaneous pancreatitis was observed in line #2 mice but not in line #1 mice. When the authors challenged the transgenic and control mice with caerulein, only the transgenic *PRSS2*^{WT} mice developed severe pancreatitis. In all three studies, the transgenic mice were shown to exhibit increased expression of trypsin in the pancreata.

In short, transgenically expressed WT human trypsinogen had a significant influence on the development of murine pancreatitis in a dosage-dependent manner (Figure 2).

Discussion

Herein, we have assessed our current knowledge of the rare trypsinogen duplication and triplication CNVs and the common rs10273639C-tagged allele. We show that there is an unambiguous positive correlation between increased trypsinogen gene dosage and CP risk in the context of the rare duplication and triplication CNVs, and between the level of

trypsinogen expression and disease risk in the context of the heterozygous and homozygous rs10273639C-tagged genotypes. We have attempted to quantify the genetic effects associated with the four genotypes involved, thereby providing a first glimpse of the impact of "quantitative" trypsinogen variants on disease risk. As far as the currently available data are concerned, the triplication genotype and the heterozygous rs10273639C-tagged genotype represent two extremes of *PRSS1*-related genotypes in terms of their genetic effect sizes, the former being causative for Mendelian HCP, the latter conferring only a modest risk for CP (Figure 1).

At this juncture, it is pertinent to make three points. First, the presence of rare known CP-associated *PRSS1* missense variants was excluded in all carriers of the published trypsinogen duplication and triplication CNVs.[30, 31, 33, 45, 46] By contrast, the status of the common rs10273639C variant was not known in these carriers. However, any effect emanating from the common rs10273639C-tagged variant would have been dwarfed by that of the rare duplication or triplication variant if and when the rs10273639C variant co-occurred with a trypsinogen gene duplication/triplication. Second, three interrelated factors, namely rs10273639C as a common variant, a minor and dosage-dependent effect of the rs10273639C-tagged allele on trypsinogen expression in human pancreatic tissue, and disease association data from multiple populations, renders the positive correlation between increased trypsinogen dosage and genetic effect robust. By contrast, the correlation between gene dosage and genetic effect in the context of rare trypsinogen copy number gains should be regarded as provisional given the relatively limited genetic data on the one hand and a possible threshold effect on the other. Third, the trypsinogen triplication CNV, which was initially identified in HCP families, has now also been found in FCP and ICP patients (Table 1). For comparison, p.Arg122His and p.Asn29Ile, the first and second variants reported in the *PRSS1* gene, were both initially found in HCP families;[11, 50] they were also found in FCP and

ICP patients and even in healthy subjects in subsequent studies (e.g., references[58-60]). Indeed, in dominantly inherited human diseases, incomplete penetrance tends to be the rule rather than the exception.[61]

As we show above, the genetic evidence to support the contention that increased expression of the WT trypsinogen protein (either through a trypsinogen copy number gain or via the common rs10273639C-tagged allele) predisposes to, or even causes CP, is convincing. However, unlike the situation with *PRSS1* missense variants, direct experimental evidence from mouse models to confirm the pathogenicity of these variants has hitherto been lacking. We have however retrospectively uncovered such evidence from three studies that were indirectly informative for the effect of transgenically expressed WT human *PRSS1* or *PRSS2* on the development or severity of pancreatitis as compared to control mice.[25, 29, 55] In this regard, we would like to emphasize two points. First, for a variety of reasons, findings from mouse CP models may not always be extrapolable to human clinical and genetic studies.[62] Second, *PRSS1*^{WT} and *PRSS2*^{WT} mice were regarded as being functionally equivalent here, providing that both of them transgenically expressed a human *trypsinogen*. Keeping these points in mind, a rough correlation between transgene expression and the development or severity of pancreatitis was apparent in all three studies.[25, 29, 55] Although the detailed downstream mechanisms remain to be elucidated, the effects of the transgene on the development or severity of pancreatitis appear to have been due to the increased level of "*human trypsin*" in the mouse pancreatic tissues.

The consistency and mutually reinforcing nature of the available human genetic and murine transgenic evidence, together with trypsinogen's inherent property of autoactivation,[63, 64] would seem to support our contention that increased expression of WT trypsinogen translates into increased production of prematurely activated trypsin within the pancreas, thereby causing or predisposing to pancreatitis depending upon the amount of

activated trypsin. The other side of the same coin is that loss-of-function *PRSS1* and *PRSS2* variants may protect against CP,[65, 66] further underscoring the importance of the level of WT trypsinogen expression in the etiology of CP. It should be noted here that CP-associated *PRSS1* missense variants, irrespective of whether or not they increase zymogen autoactivation and/or zymogen/enzyme stability, confer their pathogenic effects by increasing the level of *trypsin* within the pancreas.[27, 29] Given that *trypsin* is derived directly from trypsinogen, we propose that specifically downregulating trypsinogen expression in the pancreas could serve as a potential therapeutic and prevention strategy for CP in subjects harboring *PRSS1* risk variants. Moreover, given the fundamental role of trypsinogen in CP, specifically downregulating trypsinogen expression in the pancreas might also be a treatment option for pancreatitis patients [JIAN-MIN: do you mean specifically idiopathic CP patients?] who carry no risk variants in the *PRSS1* gene.

This study has its limitations. For example, PubMed-indexed genetic studies in the context of trypsinogen CNVs have rarely originated from outside of the Brest genetics laboratory. We hope that this study will stimulate the routine analysis of this type of variant with a high genetic effect in other laboratories .

In summary, a synthesis of our knowledge based on human genetic and murine transgenic studies significantly improves our understanding of the fundamental contributory role of the level of trypsinogen in the etiology of CP. The emergence of elevated WT trypsinogen as a key factor in increasing CP risk suggests the potential for the development of a range of new therapeutic and disease prevention strategies.

References

1. Beyer G, Habtezion A, Werner J, et al. Chronic pancreatitis. *Lancet* 2020; 396: 499-512.
2. Singh VK, Yadav D, Garg PK. Diagnosis and management of chronic pancreatitis: a review. *JAMA* 2019; 322: 2422-34.
3. Machicado JD, Amann ST, Anderson MA, et al. Quality of life in chronic pancreatitis is determined by constant pain, disability/unemployment, current smoking, and associated co-morbidities. *Am J Gastroenterol* 2017; 112: 633-42.
4. Lowenfels AB, Maisonneuve P, Cavallini G, et al. Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. *N Engl J Med* 1993; 328: 1433-7.
5. Xiao AY, Tan ML, Wu LM, et al. Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies. *Lancet Gastroenterol Hepatol* 2016; 1: 45-55.
6. Kleeff J, Whitcomb DC, Shimosegawa T, et al. Chronic pancreatitis. *Nat Rev Dis Primers* 2017; 3: 17060.
7. Hart PA, Conwell DL. Chronic pancreatitis: managing a difficult disease. *Am J Gastroenterol* 2020; 115: 49-55.
8. Le Bodic L, Bignon JD, Raguenes O, et al. The hereditary pancreatitis gene maps to long arm of chromosome 7. *Hum Mol Genet* 1996; 5: 549-54.
9. Whitcomb DC, Preston RA, Aston CE, et al. A gene for hereditary pancreatitis maps to chromosome 7q35. *Gastroenterology* 1996; 110: 1975-80.
10. Pandya A, Blanton SH, Landa B, et al. Linkage studies in a large kindred with hereditary pancreatitis confirms mapping of the gene to a 16-cM region on 7q. *Genomics* 1996; 38: 227-30.
11. Whitcomb DC, Gorry MC, Preston RA, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet* 1996; 14: 141-5.
12. Chiari H. Über die Selbstverdauung des menschlichen Pankreas. *Zeitschrift für Heilkunde* 1896; 17: 69-96.
13. Genetic Risk Factors in Chronic Pancreatitis. Available at: http://www.pancreasgenetics.org/index.php. Accessed 27 September 2021.
14. Sahin-Tóth M, Tóth M. Gain-of-function mutations associated with hereditary pancreatitis enhance autoactivation of human cationic trypsinogen. *Biochem Biophys Res Commun* 2000; 278: 286-9.
15. Sahin-Tóth M. Human cationic trypsinogen. Role of Asn-21 in zymogen activation and implications in hereditary pancreatitis. *J Biol Chem* 2000; 275: 22750-5.
16. Chen JM, Kukor Z, Le Marechal C, et al. Evolution of trypsinogen activation peptides. *Mol Biol Evol* 2003; 20: 1767-77.
17. Teich N, Le Maréchal C, Kukor Z, et al. Interaction between trypsinogen isoforms in genetically determined pancreatitis: mutation E79K in cationic trypsin (PRSS1) causes increased transactivation of anionic trypsinogen (PRSS2). *Hum Mutat* 2004; 23: 22-31.
18. Kereszturi E, Szmola R, Kukor Z, et al. Hereditary pancreatitis caused by mutation-induced misfolding of human cationic trypsinogen: a novel disease mechanism. *Hum Mutat* 2009; 30: 575-82.
19. Schnur A, Beer S, Witt H, et al. Functional effects of 13 rare *PRSS1* variants presumed to cause chronic pancreatitis. *Gut* 2014; 63: 337-43.
20. Balazs A, Hegyi P, Sahin-Tóth M. Pathogenic cellular role of the p.L104P human cationic trypsinogen variant in chronic pancreatitis. *Am J Physiol Gastrointest Liver Physiol* 2016; 310: G477-86.
21. Nemeth BC, Szucs A, Hegyi P, et al. Novel *PRSS1* mutation p.P17T validates pathogenic relevance of CTRC-mediated processing of the trypsinogen activation peptide in chronic pancreatitis. *Am J Gastroenterol* 2017; 112: 1896-98.

22. Jancso Z, Oracz G, Kujko AA, et al. Novel pathogenic *PRSS1* variant p.Glu190Lys in a case of chronic pancreatitis. *Front Genet* 2019; 10: 46.
23. Archer H, Jura N, Keller J, et al. A mouse model of hereditary pancreatitis generated by transgenic expression of R122H trypsinogen. *Gastroenterology* 2006; 131: 1844-55.
24. Selig L, Sack U, Gaiser S, et al. Characterisation of a transgenic mouse expressing R122H human cationic trypsinogen. *BMC Gastroenterol* 2006; 6: 30.
25. Athwal T, Huang W, Mukherjee R, et al. Expression of human cationic trypsinogen (PRSS1) in murine acinar cells promotes pancreatitis and apoptotic cell death. *Cell Death Dis* 2014; 5: e1165.
26. Geisz A, Sahin-Tóth M. A preclinical model of chronic pancreatitis driven by trypsinogen autoactivation. *Nat Commun* 2018; 9: 5033.
27. Jancso Z, Sahin-Tóth M. Mutation that promotes activation of trypsinogen increases severity of secretagogue-induced pancreatitis in mice. *Gastroenterology* 2020; 158: 1083-94.
28. Gui F, Zhang Y, Wan J, et al. Trypsin activity governs increased susceptibility to pancreatitis in mice expressing human PRSS1^{R122H}. *J Clin Invest* 2020; 130: 189-202.
29. Huang H, Swidnicka-Siergiejko AK, Daniluk J, et al. Transgenic expression of PRSS1^{R122H} sensitizes mice to pancreatitis. *Gastroenterology* 2020; 158: 1072-82 e7.
30. Le Maréchal C, Masson E, Chen JM, et al. Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* 2006; 38: 1372-4.
31. Masson E, Le Maréchal C, Chandak GR, et al. Trypsinogen copy number mutations in patients with idiopathic chronic pancreatitis. *Clin Gastroenterol Hepatol* 2008; 6: 82-8.
32. Whitcomb DC, LaRusch J, Krasinskas AM, et al. Common genetic variants in the *CLDN2* and *PRSS1-PRSS2* loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet* 2012; 44: 1349-54.
33. Masson E, Chen JM, Cooper DN, et al. *PRSS1* copy number variants and promoter polymorphisms in pancreatitis: common pathogenetic mechanism, different genetic effects. *Gut* 2018; 67: 592-93.
34. Herzig AF, Genin E, Cooper DN, et al. Role of the common *PRSS1-PRSS2* haplotype in alcoholic and non-alcoholic chronic pancreatitis: Meta- and re-analyses. *Genes (Basel)* 2020; 11: 1349.
35. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283: 2008-12.
36. ClinVar. Available at: https://www.ncbi.nlm.nih.gov/clinvar/. Accessed 27 September 2021.
37. Masson E, Chen JM, Scotet V, et al. Association of rare chymotrypsinogen C (*CTRC*) gene variations in patients with idiopathic chronic pancreatitis. *Hum Genet* 2008; 123: 83-91.
38. Chen JM, Férec C. Chronic pancreatitis: genetics and pathogenesis. *Annu Rev Genomics Hum Genet* 2009; 10: 63-87.
39. Chen JM, Herzig AF, Genin E, et al. Scale and scope of gene-alcohol interactions in chronic pancreatitis: a systematic review. *Genes (Basel)* 2021; 12.
40. gnomAD (Genome Aggregation Database). Available at: https://gnomad.broadinstitute.org/. Accessed 27 September 2021.
41. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-53.
42. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre TCC, 2014. https://review-manager.software.informer.com/5.3/.
43. GIGA calculator. Available at: https://www.gigacalculator.com/calculators/odds-ratio-calculator.php. Accessed 27 Septmeber 2021.
44. Chauvin A, Chen JM, Quemener S, et al. Elucidation of the complex structure and origin of the human trypsinogen locus triplication. *Hum Mol Genet* 2009; 18: 3605-14.

45. LaRusch J, Barmada MM, Solomon S, et al. Whole exome sequencing identifies multiple, complex etiologies in an idiopathic hereditary pancreatitis kindred. *JOP* 2012; 13: 258-62.
46. Wang W, Sun XT, Weng XL, et al. Comprehensive screening for *PRSS1*, *SPINK1*, *CFTR*, *CTRC* and *CLDN2* gene mutations in Chinese paediatric patients with idiopathic chronic pancreatitis: a cohort study. *BMJ Open* 2013; 3: e003150.
47. Masson E, Le Maréchal C, Delcenserie R, et al. Hereditary pancreatitis caused by a double gain-of-function trypsinogen mutation. *Hum Genet* 2008; 123: 521-9.
48. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature* 2020; 581: 444-51.
49. Le Maréchal C, Chen JM, Le Gall C, et al. Two novel severe mutations in the pancreatic secretory trypsin inhibitor gene (*SPINK1*) cause familial and/or hereditary pancreatitis. *Hum Mutat* 2004; 23: 205.
50. Gorry MC, Gabbaizedeh D, Furey W, et al. Mutations in the cationic trypsinogen gene are associated with recurrent acute and chronic pancreatitis. *Gastroenterology* 1997; 113: 1063-8.
51. Boulling A, Sato M, Masson E, et al. Identification of a functional *PRSS1* promoter variant in linkage disequilibrium with the chronic pancreatitis-protecting rs10273639. *Gut* 2015; 64: 1837-8.
52. Derikx MH, Kovacs P, Scholz M, et al. Polymorphisms at *PRSS1-PRSS2* and *CLDN2-MORC4* loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. *Gut* 2015; 64: 1426-33.
53. Masamune A, Nakano E, Hamada S, et al. Common variants at *PRSS1-PRSS2* and *CLDN2-MORC4* loci associate with chronic pancreatitis in Japan. *Gut* 2015; 64: 1345-6.
54. Paliwal S, Bhaskar S, Nageshwar Reddy D, et al. Association analysis of *PRSS1-PRSS2* and *CLDN2-MORC4* variants in nonalcoholic chronic pancreatitis using tropical calcific pancreatitis as model. *Pancreas* 2016; 45: 1153-7.
55. Wan J, Haddock A, Edenfield B, et al. Transgenic expression of human *PRSS2* exacerbates pancreatitis in mice. *Gut* 2020; 69: 2051-52.
56. Chen JM, Férec C. Genes, cloned cDNAs, and proteins of human trypsinogens and pancreatitis-associated cationic trypsinogen mutations. *Pancreas* 2000; 21: 57-62.
57. Kukor Z, Toth M, Sahin-Tóth M. Human anionic trypsinogen: properties of autocatalytic activation and degradation and implications in pancreatic diseases. *Eur J Biochem* 2003; 270: 2047-58.
58. Khalid A, Finkelstein S, Thompson B, et al. A 93 year old man with the *PRSS1* R122H mutation, low *SPINK1* expression, and no pancreatitis: insights into phenotypic non-penetrance. *Gut* 2006; 55: 728-31.
59. Masson E, Chen JM, Audrézet MP, et al. 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* 2013; 8: e73522.
60. Zou WB, Tang XY, Zhou DZ, et al. *SPINK1*, *PRSS1*, *CTRC*, and *CFTR* genotypes influence disease onset and clinical outcomes in chronic pancreatitis. *Clin Transl Gastroenterol* 2018; 9: 204.
61. Cooper DN, Krawczak M, Polychronakos C, et al. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* 2013; 132: 1077-130.
62. Sendler M, Lerch MM. The complex role of trypsin in pancreatitis. *Gastroenterology* 2020; 158: 822-26.
63. Kassell B, Kay J. Zymogens of proteolytic enzymes. *Science* 1973; 180: 1022-7.
64. Chen JM, Montier T, Férec C. Molecular pathology and evolutionary and physiological implications of pancreatitis-associated cationic trypsinogen mutations. *Hum Genet* 2001; 109: 245-52.

65. Chen JM, Le Maréchal C, Lucas D, et al. "Loss of function" mutations in the cationic trypsinogen gene (*PRSS1*) may act as a protective factor against pancreatitis. *Mol Genet Metab* 2003; 79: 67-70.
66. Witt H, Sahin-Tóth M, Landt O, et al. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. *Nat Genet* 2006; 38: 668-73.

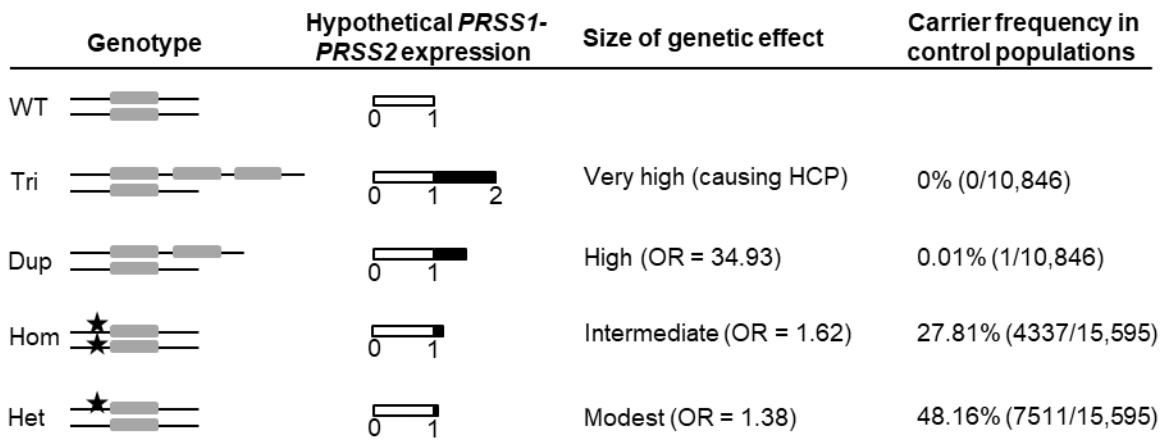


Figure 1. Illustration of the trypsinogen (*PRSS1* and *PRSS2*) copy number gains and rs10273639C-tagged genotypes, their anticipated hypothetical trypsinogen expression levels, genetic effect sizes and global carrier frequencies. It should be noted that the *PRSS1-PRSS2* expression levels in carriers of trypsinogen copy number gain variants are purely hypothetical whereas those in carriers of rs10273639C-tagged heterozygous and homozygous genotypes were supported by experimental data. See text for details about how the genetic effect sizes were determined. Carrier frequencies in control populations were derived from gnomAD (SVs v2.1 for the Tri and Dup genotypes and v2.1.1 for the rs10273639C-tagged genotypes).[40] Grey bar, the *PRSS1-PRSS2* loci on chromosome 7q34. Star, rs10273639C allele. WT, wild-type; Tri, triplication; Dup, duplication; Hom, homozygote; Het, heterozygote; HCP, hereditary chronic pancreatitis; OR, odds ratio.

	Mice	Transgene expression	Spontaneous pancreatitis	Pancreatitis after caerulein challenge
Athwal et al.	Control	No	No	Yes
	PRSS1^{WT}	Yes	Yes (occurred in up to 10% of ageing (>9 month) animals)	More severe
Huang et al.	Control	No	No	Minimal
	PRSS1^{WT}	Yes	Yes (focal; in homozygous mice)	Severe
Wan et al.	Control	No	No	Minimal
	PRSS2^{WT} line #1	Yes ↑	No	Severe
	PRSS2^{WT} line #2	Yes ↑↑	Yes (focal; area < 10%)	Severe

Figure 2. Summary of findings from three studies (Athwal et al.[25], Huang et al.[29] and Wan et al.[55]) that reported the transgenic expression of the wild-type human *PRSS2* or *PRSS1* gene in mice. Unless specifically stated, the zygosity of the transgene in the genetically modified mice was unknown. In the Wan et al. study, the number of upward pointing arrows indicates the different levels of the expressed *PRSS2* gene in the two mouse lines. *PRSS2*^{WT}, wild-type *PRSS2* gene; *PRSS1*^{WT}, wild-type *PRSS1* gene.

Table 1. PubMed-indexed genetic studies of trypsinogen copy number gain variants included in the current analysis

Study	Copy number gain	Disease subtype (number of families or patients affected)
Le Maréchal et al. (2006)[30]	Triplication[a,b]	HCP (5 French Caucasian families)
Masson et al. (2008)[31]	Triplication (same as above)	ICP (10 French Caucasian patients)
Masson et al. (2008)[31]	Duplication (#1)[a,b]	ICP (4 French Caucasian patients)
Masson et al. (2018)[33]	Duplication (#2)[b]	FCP (1 French Caucasian patient)
	Duplication (#3)[b]	ICP (1 Maghrebian patient)
	Duplication (#4)[b]	ICP (1 French Caucasian patient)
	Duplication (#5)[b]	ICP (1 French Guyanese patient)
LaRush et al. (2012)[45]	Duplication	Detected only in the youngest patient from an HCP family studied in USA; it must have originated *de novo* since it was not detected in either parent
Wang et al. (2013)[46]	Duplication	ICP (1 Chinese patient)
Masson et al. (2008)[47]	Double gain-of-function duplication[b]	HCP (1 French Caucasian family)

[a]Originated from a common founder chromosome.[31, 44]

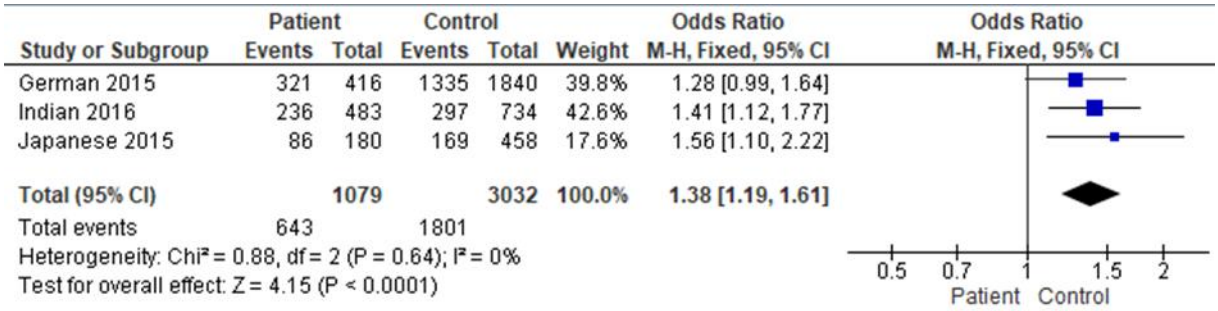
[b]Breakpoints characterized at the nucleotide sequence level.

Abbreviations: FCP, familial chronic pancreatic; HCP, hereditary chronic pancreatitis; ICP, idiopathic chronic pancreatitis.

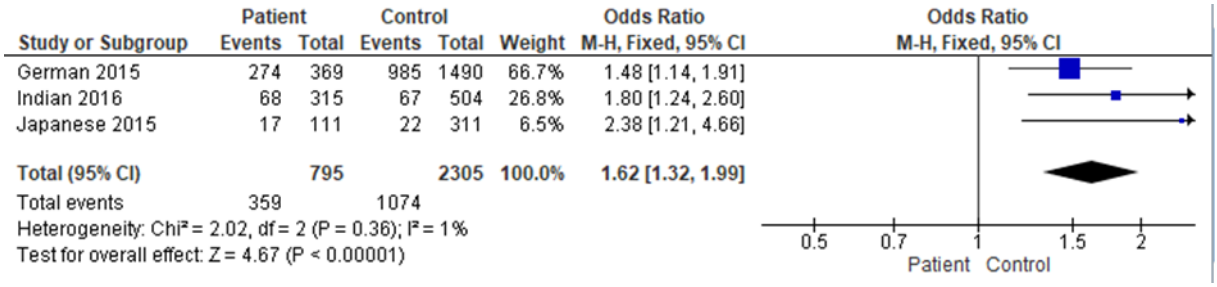
Table 2. ClinVar-registered *PRSS1* copy number gain variants that have been described to (i) involve all five exons of the *PRSS1* gene and (ii) to be associated with "hereditary pancreatitis" (as of 27 September 2021)

ClinVar Accession number	Submitter	Duplication or triplication	Ethnicity of the proband[a]	
				Disease subtype[b]
VCV000584097.3	Invitae	Triplication	Hispanic (#11)	FCP
		Triplication	Hispanic (#17)	HCP
		Duplication	Hispanic (#4)	ICP
		Duplication	White/Caucasian (#22)	ICP
VCV000417552.1	Invitae	Duplication	White/Caucasian (#16)	ICP (proband also carried *PRSS1* p.Arg122His; mother, sister and daughter carried the duplication but were not known to be clinically affected)
VCV000267330.1	Invitae	Duplication	Hispanic (#19)	FCP
VCV000973568.1	Centogene AG	Duplication	—[c]	—
VCV000831511.1	Invitae	Triplication	White/Caucasian (#9)	FCP
		Triplication	White/Caucasian (#10)	ICP
		Triplication	Hispanic (#12)	FCP
		Duplication	Hispanic (#5)	ICP
		Duplication	Unknown (#24)	ICP
VCV000831250.1	Invitae	Duplication	White/Caucasian (#14)	ICP

[a]None of the probands were known to be related.
[b]Redefined as described in Methods.
[c]No reply from the submitter.



Supplementary Figure S1. Meta-analysis of the odds ratios for the association between the heterozygous s10273639C variant and pancreatitis risk. Events, number of heterozygous rs10273639C carriers. Total, number of heterozygous rs10273639C carriers plus number of homozygous rs10273639T carriers. German study, Derikx et al.[52]; Indian study, Paliwal et al.[54]; Japan study, Masamune et al.[53] M-H, Mantel-Haenszel. Meta-analysis was performed by means of the Review Manager 5.3 software,[42] as previously described.[34]



Supplementary Figure S2. Meta-analysis of the odds ratios for the association between the homozygous rs10273639C variant and pancreatitis risk. Events, number of homozygous rs10273639C carriers. Total, number of homozygous rs10273639C carriers plus number of homozygous rs10273639T carriers. German study, Derikx et al.[52]; Indian study, Paliwal et al.[54]; Japan study, Masamune et al.[53] M-H, Mantel-Haenszel. Meta-analysis was performed by means of the Review Manager 5.3 software,[42] as previously described.[34]