nature genetics

A degradation-sensitive anionic trypsinogen (*PRSS2*) variant protects against chronic pancreatitis

Heiko Witt[1,2,39], Miklós Sahin-Tóth[3,39], Olfert Landt[4], Jian-Min Chen[5], Thilo Kähne[6], Joost PH Drenth[7], Zoltán Kukor[3], Edit Szepessy[3], Walter Halangk[8], Stefan Dahm[9], Klaus Rohde[9], Hans-Ulrich Schulz[8], Cédric Le Maréchal[5], Nejat Akar[10], Rudolf W Ammann[11], Kaspar Truninger[11,12], Mario Bargetzi[13], Eesh Bhatia[14], Carlo Castellani[15], Giulia Martina Cavestro[16], Milos Cerny[17], Giovanni Destro-Bisol[18], Gabriella Spedini[18], Hans Eiberg[19], Jan B M J Jansen[7], Monika Koudova[20], Eva Rausova[20], Milan Macek Jr[20], Núria Malats[21], Francisco X Real[21], Hans-Jürgen Menzel[22], Pedro Moral[23], Roberta Galavotti[24], Pier Franco Pignatti[24], Olga Rickards[25], Julius Spicak[26], Narcis Octavian Zarnescu[27], Wolfgang Böck[28], Thomas M Gress[28], Helmut Friess[29], Johann Ockenga[30], Hartmut Schmidt[30,31], Roland Pfützer[32], Matthias Löhr[32], Peter Simon[33], Frank Ulrich Weiss[33], Markus M Lerch[33], Niels Teich[34], Volker Keim[34], Thomas Berg[1], Bertram Wiedenmann[1], Werner Luck[2], David Alexander Groneberg[2], Michael Becker[35], Thomas Keil[36], Andreas Kage[37], Jana Bernardova[1,2], Markus Braun[1,2], Claudia Güldner[1,2], Juliane Halangk[1], Jonas Rosendahl[2,34], Ulrike Witt[38], Matthias Treiber[1,2], Renate Nickel[2] & Claude Férec[5]

Chronic pancreatitis is a common inflammatory disease of the pancreas. Mutations in the genes encoding cationic trypsinogen (*PRSS1*)[1] and the pancreatic secretory trypsin inhibitor (*SPINK1*)[2] are associated with chronic pancreatitis. Because increased proteolytic activity owing to mutated *PRSS1* enhances the risk for chronic pancreatitis, mutations in the gene encoding anionic trypsinogen (*PRSS2*) may also predispose to disease. Here we analyzed *PRSS2* in individuals with chronic pancreatitis and controls and found, to our surprise, that a variant of codon 191 (G191R) is overrepresented in control subjects: G191R was present in 220/6,459 (3.4%) controls but in only 32/2,466 (1.3%) affected individuals (odds ratio 0.37; $P = 1.1 \times 10^{-8}$). Upon activation by enterokinase or trypsin, purified recombinant G191R protein showed a complete loss of trypsin activity owing to the introduction of a new tryptic cleavage site that renders the enzyme hypersensitive to autocatalytic proteolysis. In conclusion, the G191R variant of PRSS2 mitigates intrapancreatic trypsin activity and thereby protects against chronic pancreatitis.

[1]Department of Hepatology and Gastroenterology, Charité University Hospital, Augustenburger Platz 1, 13353 Berlin, Germany. [2]Department of Pediatrics, Charité University Hospital, Augustenburger Platz 1, 13353 Berlin, Germany. [3]Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA. [4]TIB MOLBIOL, Eresburgstrasse 22-23, 12103 Berlin, Germany. [5]Institut National de la Santé et de la Recherche Médicale (INSERM), U613, Brest, F-29220 France; Université de Bretagne Occidentale, Faculté de Médecine de Brest et des Sciences de la Santé, F-29238 France; Etablissement Français du Sang-Bretagne, Brest and Centre Hospitalier Régional Universitaire (CHRU) Brest, Hôpital Morvan, Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Brest, F-29220 France. [6]Institute of Experimental Internal Medicine, Otto-von-Guericke University, Leipziger Strasse 44, 39120 Magdeburg, Germany. [7]Department of Medicine, Division of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Center, 6500 HB Nijmegen, The Netherlands. [8]Department of Surgery, Otto-von-Guericke University, Leipziger Strasse 44, 39120 Magdeburg, Germany. [9]Department of Bioinformatics, Max-Delbrück-Centrum, 13092 Berlin, Germany. [10]Department of Pediatric Molecular Genetics, Ankara University Medical School, Konutkent-2, Mudanya Sokak C-1 Blok B-2, 06530 Cayyolu, Ankara, Turkey. [11]Department of Medicine, Division of Gastroenterology, University Hospital, Rämistrasse 100, 8092 Zurich, Switzerland. [12]Department of Medicine, Division of Gastroenterology, Kantonsspital Aarau, Tellstrasse, CH-5001 Aarau, Switzerland. [13]Department of Medicine, Center of Oncology/Hematology, Kantonsspital Aarau, Tellstrasse, CH-5001 Aarau, Switzerland. [14]Department of Endocrinology, Sanjay Gandhi Postgraduate Institute, Lucknow 226014, India. [15]Cystic Fibrosis Centre, Ospedale Civile Maggiore, Piazzale Stefani 1, 37126 Verona, Italy. [16]Gastroenterology Section, Department of Clinical Sciences, University of Parma, Via Gramsci 14, 43100 Parma, Italy. [17]Clinic of Obstetrics and Gynecology, Neonatology Department, University Hospital Motol and Second Medical School, Charles University Prague, V Uvalu 84, 103 Prague 5, Czech Republic. [18]Department of Human and Animal Biology, Section of Anthropology, University of Rome "La Sapienza", Via della Ricerca Scientifica 1, 00133 Rome, Italy. [19]Department of Medical Biochemistry and Genetics, Panum Institute, University of Copenhagen, IMBG-G Build 24.4, Blegdamsvej 3, DK 2200, Copenhagen, Denmark. [20]Institute of Biology and Medical Genetics–Cystic Fibrosis Center, University Hospital Motol and Second School of Medicine of Charles University, V Uvalu 84, Prague 5, CZ 150 06, Czech Republic. [21]Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Carrer Dr. Aiguader 80, 08003 Barcelona, Spain. [22]Institute of Medical Biology and Human Genetics, Medical University of Innsbruck, Schöpfstr. 41, 6020 Innsbruck, Austria. [23]Unitat d'Antropologia, Departament de Biologia Animal, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain. [24]Department Maternal Infantile and of Biology-Genetics (DMIBG), Section of Biology and Genetics, University of Verona, Strada le Grazie 37134 Verona, Italy. [25]Department of Biology, University of Rome "Tor Vergata", Piazzale A. Moro 5, 00185 Rome, Italy. [26]Clinic of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (IKEM), Videnska 1958/9, 14021 Prague 4, Czech Republic. [27]Second Department of Surgery, University Emergency Hospital, 72435 Bucharest, Romania. [28]Department of Internal Medicine I, University Hospital Ulm, Robert Koch Str. 8, 89081 Ulm, Germany. [29]Department of General Surgery, University of Heidelberg, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany. [30]Department of Gastroenterology, Hepatology & Endocrinology, Charité University Hospital, 10008 Berlin, Germany. [31]Transplant Hepatology, University Hospital Münster, Albert-Schweitzer-Strasse 33 48149 Münster, Germany. [32]Department of Medicine II (Gastroenterology/Hepatology/Infectious Diseases), University of Heidelberg, Medical Faculty of Mannheim, Theodor-Kutzer-Ufer, 68167 Mannheim, Germany. [33]Department of Gastroenterology, Endocrinology and Nutrition, Ernst-Moritz-Arndt University, Friedrich-Loeffler-Str. 23A, 17487 Greifswald, Germany. [34]Department of Gastroenterology and Hepatology, University of Leipzig, Philipp-Rosenthal-Str. 27, 04103 Leipzig, Germany. [35]Department of Pediatric Gastroenterology, Deutsches Rotes Kreuz Kliniken Westend, Spandauer Damm 130, 14050 Berlin, Germany. [36]Institute for Social Medicine and Epidemiology, Charité University Hospital, 10008 Berlin, Germany. [37]Institute of Laboratory Medicine and Pathobiochemistry and [38]Department of Surgery, Charité University Hospital, Augustenburger Platz 1, 13353 Berlin, Germany. [39]These authors contributed equally to this work. Correspondence should be addressed to H.W. (heiko.witt@charite.de).

Received 28 April 2005; accepted 7 April 2006; published online 14 May 2006; doi:10.1038/ng1797

Three different isoforms of trypsinogen have been described in human pancreatic secretions. According to their electrophoretic mobility on isoelectric focusing, they have been designated as cationic trypsinogen (PRSS1), anionic trypsinogen (PRSS2) and mesotrypsinogen (PRSS3)[3]. Anionic trypsinogen, also referred to as serine protease 2, is one of the most abundant secretory proteins synthesized by the pancreas[3,4]. The ratio of the anionic to the cationic forms is approximately 1:2, whereas mesotrypsinogen is secreted only in low quantities[3]. Compared with the cationic isoenzyme, anionic trypsinogen autoactivates less easily, particularly at acidic pH, and is more sensitive to autolysis[5,6]. The gene encoding human PRSS2 (OMIM 601564) maps to 7q35, is approximately 3.6 kb long and contains five exons[7]. The preproprotein comprises 247 amino acids, including a 15-residue signal peptide and an eight-residue activation peptide.

Mutations in *PRSS1* (OMIM 276000) and *SPINK1* (OMIM 167790) are strongly associated with hereditary and idiopathic chronic pancreatitis[1,2]. However, the majority of affected individuals do not carry mutations in either of these two genes, suggesting that defects in further genes might be involved in the pathogenesis of the disease[2]. Because mutations in *PRSS1* can cause pancreatitis, we hypothesized that, by analogy, gene variants of the anionic isoenzyme *PRSS2* could also modify the susceptibility to chronic pancreatitis. We therefore investigated the coding regions and flanking intronic sequences of *PRSS2* in 200 German affected individuals and 240 controls by SSCP analysis and direct DNA sequencing.

In these subjects, we identified three missense PRSS2 variants. Two of the variants, T8I and V118I, were equally common among individuals with chronic pancreatitis (12.5% and 5%, respectively) and controls (15% and 4.6%, respectively). In contrast, the third variant, a G→A transition resulting in a substitution of glycine by arginine at codon 191 in exon 4 (G191R), was present in 9/240 (3.8%) control subjects but not in any of the 200 affected individuals investigated.

We subsequently expanded the study to include additional German affected individuals and controls. Using melting curve analysis, we confirmed the higher prevalence of the heterozygous G191R variant in the control group (71/2,019 (3.5%)) versus individuals with chronic pancreatitis (19/1,548 (1.2%)) (odds ratio (OR) 0.34; $P = 0.000012$) (**Table 1**).

In order to investigate whether our findings were applicable to other European populations, we investigated an additional 918 individuals with chronic pancreatitis and 4,440 control subjects from Austria, the Czech Republic, Denmark, France, Italy, the Netherlands, Romania, Spain and Switzerland. Because of the low case numbers in some of these populations, we pooled them together as a European sample without Germans. The frequency (3.4%) and 95% confidence interval (c.i.) (2.8–3.9) for the controls in this European sample lay well in the range of the German sample (frequency = 3.5%; 95% c.i. = 2.7–4.3). The frequency in controls (149/4,440 (3.4%)) was significantly higher than in individuals with chronic pancreatitis (13/918 (1.4%)) (OR 0.41; $P = 0.00097$) (**Table 1**). Taking all sampled European populations together, the frequency of G191R was 1.3% (32/2,466) in the group of affected individuals, compared with 3.4% (220/6,459) among control subjects (OR = 0.37; $P = 1.1 \times 10^{-8}$) (**Table 1**). Stratification of the subgroups (alcoholic pancreatitis and idiopathic/hereditary pancreatitis) showed no significant differences between the different groups of affected individuals. For detailed data, see **Supplementary Table 1** online.

Further analyses showed that affected individuals with the G191R variant were of higher age than those without the protective variant. In the idiopathic/hereditary pancreatitis group, we found G191R in 24/1,256 (1.9%) affected individuals older than 20 years compared with 3/601 (0.5%) affected individuals aged 20 years or younger ($P = 0.021$); however, owing to limited sample size, these data should be interpreted with caution. A similar tendency was observed in the alcoholic chronic pancreatitis group: none of the 162 affected individuals aged 40 years or younger had the G191R variant, compared with 5/447 (1.1%) of affected individuals older than 40 years ($P = 0.332$; **Supplementary Table 1**). No age-related differences were observed in the control group (3.4% in individuals aged 20 years or younger versus 3.3% in those over 20 years; 3.2% in individuals aged 40 years or younger versus 3.5% in individuals over 40 years).

In contrast to unaffected Europeans, who showed a G191R frequency of 3.4%, the variant was rare in individuals of African descent (2/948), with an estimated overall frequency of 0.2% (95% c.i. = 0–0.5), which was significantly different from the frequency we found in the European population ($P = 10^{-11}$).

One of the 6,459 control subjects (originating from the Netherlands), but none of the affected individuals, was homozygous for G191R. When analyzing the G191R variation among 204 parents of Austrian or German individuals with chronic pancreatitis, we detected six G191R carriers. However, the variant was not transmitted from the unaffected parent to the affected child in any of the cases.

To exclude that G191R is in linkage disequilibrium (LD) with other variants in *PRSS2* or *PRSS1*, which is located only 40 kb centromeric to *PRSS2*, we bidirectionally sequenced the complete *PRSS1* and *PRSS2* coding sequence in 60 *PRSS2* G191R heterozygotes (ten affected individuals and 50 controls). With exception of the common *PRSS2* polymorphisms T8I and V118I, which were present in a similar frequency as found in the above-mentioned German affected individuals and control subjects (15% and 1.7%, respectively), we did not find any other *PRSS2* variant within the coding region. Moreover, none of these 60 G191R heterozygotes carried a *PRSS1* missense mutation.

In order to determine the haplotype structure in the genomic context of G191R, we analyzed 17 SNPs (going from *rs2156965* to *rs985180*) around *PRSS2* (**Fig. 1**). From a cohort of 95 subjects corresponding to 36 families with at least a carrier of G191R,

Table 1 Frequency of G191R *PRSS2* variant in affected individuals and controls

	Affected individuals			Controls				
Population	Number of mutations/total sample size	%	95% c.i.	Number of mutations/total sample size	%	95% c.i.	OR	*P* value
German	19/1,548	1.2	0.7–1.8	71/2,019	3.5	2.7–4.3	0.34	0.000012
European (w/o German)	13/918	1.4	0.7–2.2	149/4,440	3.4	2.8–3.9	0.41	0.00097
European (all)	32/2,466	1.3	0.9–1.7	220/6,459	3.4	3.0–3.8	0.37	1.1×10^{-8}

P values were determined by Fisher's Exact Test.

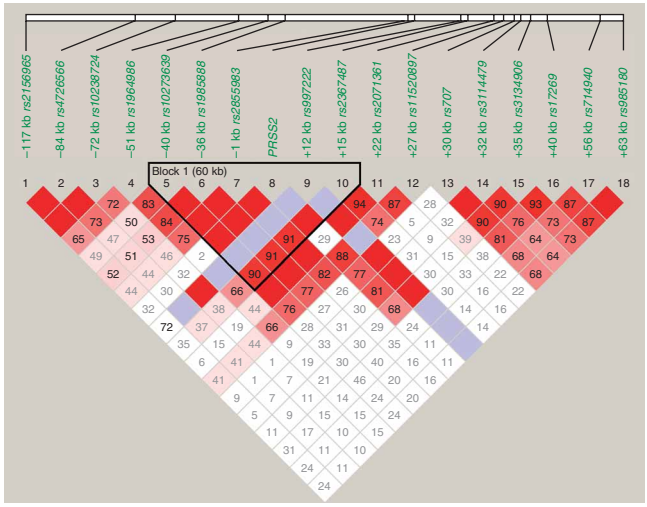

Figure 1 Haploview graph of the pairwise linkage disequilibrium (LD) among 17 SNPs around the *PRSS2* locus. Block 1 indicates the minimal haplotype always associated with *PRSS2* G191R. $D' = 1.00$ in the blank red squares; numbers inside the squares are $D' \times 100$. Genomic locations of the different SNPs are illustrated above the LD plot.

we found a small common haplotype always associated with G191R (233112 or CGGAAC in **Table 2a**; Block 1 in **Fig. 1**). This minimum haplotype starts at *rs10273639* (located within *PRSS1* 40 kb centromeric from *PRSS2*) and ends with *rs2367487* (located 15 kb telomeric from *PRSS2*) corresponding to an interval of ~60 kb ($D' = 0.90$). In contrast, we found eight different haplotypes on 72 chromosomes bearing wild-type *PRSS2* (**Table 2a**). Taking into account a SNP 51 kb centromeric from *PRSS2* (*rs1964986*) and 22 kb telomeric from *PRSS2* (*rs2071361*), we found three haplotypes: we found the first haplotype on 10/22 chromosomes (46%), the second on 8/22 chromosomes (36%) and the third on 4/22 chromosomes (18%) (**Table 2b**).

We then aimed to exclude the genetic effect of potential causative variant(s) in any non-233112 haplotypes. In all German affected individuals and control subjects, we further analyzed *rs10273639*, which tags the two most common haplotypes (233312 and 411314), and *rs997222*, which tags the only other haplotype with a significant frequency (233342; 10%) (**Table 2a**). Allele frequencies of both SNPs did not significantly differ between affected individuals and controls: *rs10273639 T* was present in 1,277/3,096 (41.2%) affected individuals versus 1,715/4,038 (42.5%) controls ($P = 0.31$), and *rs997222 T* was present in 242/3,096 (7.8%) affected individuals versus 355/4,038 (8.8%) controls ($P = 0.14$). In addition, haplotype analysis as well as the very rare frequency of G191R in individuals of African descent indicate that G191R is a founder allele and not a recurrent mutation.

The glycine at position 191 is strictly conserved in vertebrate trypsinogens. Although the crystal structure of human anionic trypsin remains to be solved, comparison with rat trypsin or the homologous human cationic trypsin (PRSS1) structure (**Fig. 2**) suggests that Gly191 is located on the surface of the enzyme near the primary substrate binding pocket. Therefore, the G191R alteration is expected to generate a surface-exposed, positively charged, bulky arginine side chain. This Arg191-Gly192 peptide bond would be predicted to be highly amenable to cleavage by trypsin. PRSS2 lacks the conserved Cys139-Cys206 disulfide bond that would prevent the dissociation of the C-terminal proteolytic fragment[8]. This fragment contains the catalytic Ser200 residue, which is essential for enzymatic activity (**Fig. 2**). The cleavage should, therefore, inactivate trypsin(ogen) rapidly and irreversibly.

To test this hypothesis, we recombinantly expressed wild-type anionic trypsinogen and the G191R variant. Both the physiological activator enteropeptidase (enterokinase) and trypsin strongly activated wild-type PRSS2, and the trypsin activity generated remained stable over the entire time period studied. In contrast, the G191R variant showed only minimal or no tryptic activity after activation (**Fig. 3**). Analysis of the activation process by reducing SDS-PAGE showed a time-dependent conversion of trypsinogen to trypsin for wild-type PRSS2. In contrast, activation of the G191R variant led to a strong

Table 2 Haplotype analysis of the G191R region

(**a**) Common haplotypes associated with G191R

rs10273639	rs1985888	rs2855983	G191R	rs997222	rs2367487	*n*	%
−40 kb	−36 kb	−1 kb	*PRSS2*	+12 kb	+15 kb		
2	3	3	1	1	2	27	100%
2	1	1	3	1	4	2	3%
2	3	1	3	1	4	2	3%
2	3	3	3	1	2	29	40%
2	3	3	3	1	4	1	1%
2	3	3	3	4	2	7	10%
2	3	3	3	4	4	1	1%
4	1	1	3	1	2	1	1%
4	1	1	3	1	4	29	40%

(**b**) Extended haplotype analysis of G191R mutant alleles

rs1964986	rs10273639	rs1985888	rs2855983	G191R	rs997222	rs2367487	rs2071361	*n*	%
−51 kb	−40 kb	−36 kb	−1 kb	***PRSS2***	+12 kb	+15 kb	+22 kb		
1	2	3	3	1	1	2	3	10	46%
2	2	3	3	1	1	2	2	8	36%
2	2	3	3	1	1	2	3	4	18%

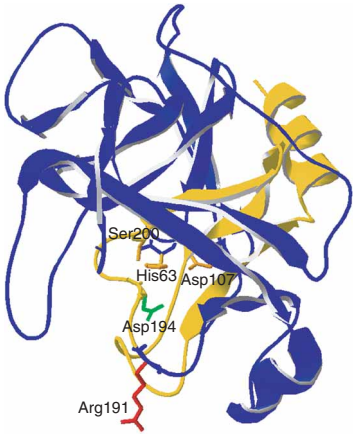

Figure 2 Ribbon diagram of human cationic trypsin (Protein Data Bank file 1TRN) showing Gly191 (chymotrypsin numbering (chymo#) Gly187) mutated to Arg (in red). Also shown is the primary specificity determinant Asp194 (in *green*; chymo#Asp189) and the catalytic triad His63, Asp107 and Ser200 (in *orange*, chymo# His57, Asp102 and Ser195). The *yellow* peptide segment is removed after cleavage of the Arg191-Gly192 peptide bond.

new band corresponding to the N-terminal proteolytic fragment of trypsinogen cleaved after Arg191 (**Fig. 3a–c**). Removal of the activation peptide from this proteolytic fragment by enteropeptidase yielded a smaller fragment corresponding to the N-terminal fragment of trypsin cleaved after Arg191 (**Fig. 3a**). We further characterized the proteolytic processing of the G191R mutant by mass spectrometry (**Fig. 4**). After incubation with enteropeptidase, wild-type trypsinogen (25,069.2 Da) was converted to trypsin (24,034 Da), whereas the G191R mutant proenzyme (25,168.3 Da) was rapidly converted to cleavage products of molecular masses corresponding to the N-terminal peptide 16–191 (19,233.7 Da) and the C-terminal peptide 192–247 (data not shown). Intact G191R trypsin was nearly undetectable. Enteropeptidase also converted the N-terminal G191R trypsinogen fragment to its corresponding trypsin fragment (18,196.5 Da) (**Fig. 4**). Incubation with trypsin as activating enzyme similarly demonstrated rapid cleavage of the G191R mutant at the Arg191–Gly192 peptide bond (data not shown). Taken together, this functional analysis confirms that the G191R substitution generates a supersensitive tryptic cleavage site and destines anionic trypsin(ogen) for autocatalytic degradation.

It is currently thought that pancreatitis results from an imbalance of proteases and their inhibitors within the pancreatic parenchyma. It has been shown that a gain-of-function mutation in the cationic trypsinogen gene is a cause of hereditary pancreatitis[1]. A more recent study provided evidence that loss-of-function mutations in the specific trypsin inhibitor, *SPINK1*, can also predispose to chronic pancreatitis[2]. Thus, it is now generally believed that inappropriate intrapancreatic activation of trypsinogen is a precondition for acinar cell injury and pancreatitis. Along the same line of thought, it is conceivable that loss-of-function trypsinogen mutations should afford protection against pancreatitis. The present study is the first to experimentally support this notion not only by identifying a previously unknown *PRSS2* variant that is significantly underrepresented in individuals with pancreatitis from various geographic regions but also by providing experimental evidence that this variant conveys a loss of function. Although the overall contribution of G191R to disease pathogenesis is low, the functional characterization of the G191R variant reported here provides the first example in pancreatitis—and a rare example as far as inheritable diseases in general are concerned—for a disease-protective genetic variant.

METHODS

Individuals affected with chronic pancreatitis. This study was approved by the medical ethical review committee of the Charité University Hospital. All affected individuals gave informed consent for genetic analysis. The study population included 2,466 individuals suffering from chronic pancreatitis (609 with alcohol-related disease and 1,857 with either idiopathic or hereditary etiology) originating from Austria (n = 24), the Czech Republic (n = 92), France (n = 359), Germany (n = 1,548), Italy (n = 141), the Netherlands (n = 139), Romania (n = 3), Spain (n = 64) and Switzerland (n = 96). The clinical diagnosis of chronic pancreatitis was based on two or more of the

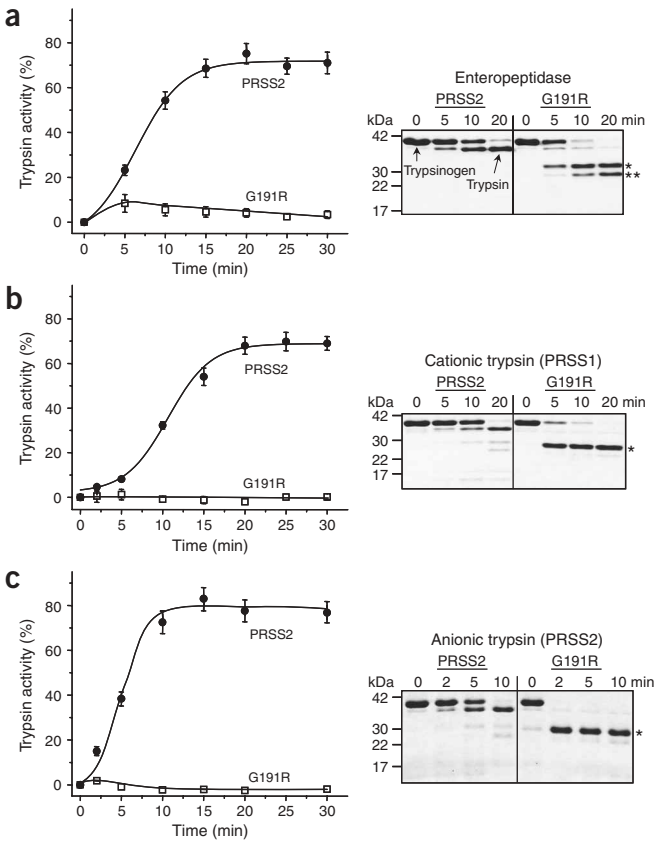

Figure 3 Activation of wild-type and Arg191 mutant anionic trypsinogens. (**a**) Enteropeptidase. (**b**) Human cationic trypsin. (**c**) Human anionic trypsin. Activation assays and gel electrophoresis were carried out as described in Methods. Trypsin activity was expressed as percentage of the potential total activity. In experiments **b** and **c**, the activity of the added trypsin was subtracted from the total activity measured. The single asterisk indicates the N-terminal trypsinogen fragment generated by cleavage of the Arg191-Gly192 peptide bond. This trypsinogen fragment was further processed by enteropeptidase to the corresponding trypsin fragment (**). The C-terminal peptide was not detected under these electrophoretic conditions. Control experiments using human pancreatic trypsin inhibitor demonstrate that enteropeptidase cannot cleave the Arg191-Gly192 peptide bond and that the degradation observed during enteropeptidase activation is entirely attributable to trypsin (data not shown).

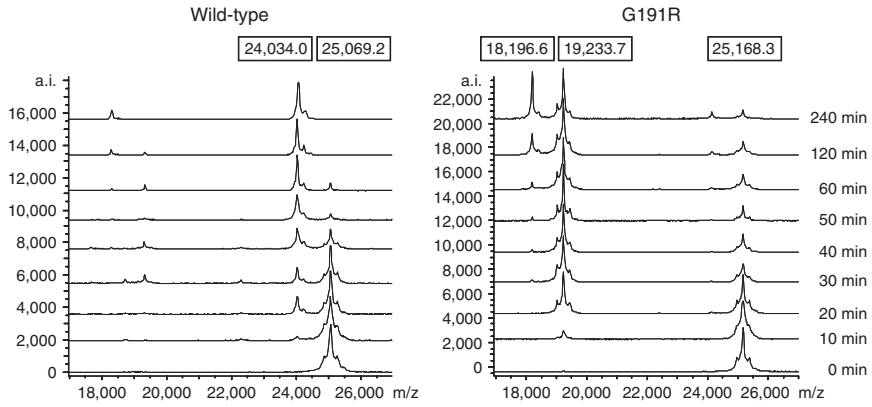

Figure 4 Mass spectrometric analysis of trypsinogen activation. Wild-type and Arg191 variant anionic trypsinogen were incubated at pH 8.0 in the presence of 10 mM $CaCl_2$, as described in Methods. At the indicated time points, samples were withdrawn and directly analyzed by MALDI-TOF mass spectrometry. The left shoulders observed on the trypsinogen peaks are due to partial N-terminal processing of recombinant trypsinogen by *Escherichia coli* aminopeptidases; the right shoulders are sinapinic acid adducts. The molecular masses indicated correspond to average peptide masses with cysteines oxidized as disulfides. Note that recombinant trypsinogens contain an extra methionine at the N terminus.

following criteria: presence of a typical history of recurrent pancreatitis, radiological findings such as pancreatic calcifications and/or pancreatic ductal irregularities revealed by endoscopic retrograde pancreatography or by magnetic resonance imaging of the pancreas and/or pathological sonographic findings. Hereditary chronic pancreatitis was diagnosed when two first-degree relatives or three or more second-degree relatives suffered from recurrent acute or chronic pancreatitis without any apparent precipitating factor. Affected individuals were classified as having idiopathic chronic pancreatitis when precipitating factors, such as alcohol abuse, trauma, medication, infection, metabolic disorders and/or a positive family history, were absent. Alcohol-induced chronic pancreatitis was diagnosed in affected individuals who consumed more than 60 g (females) or 80 g (males) of ethanol per day for more than two years.

The 6,459 control subjects were recruited from Austria ($n = 637$), the Czech Republic ($n = 522$), Denmark ($n = 95$), France ($n = 1,120$), Germany ($n = 2,019$), Italy ($n = 326$), the Netherlands ($n = 565$), Romania ($n = 163$), Spain ($n = 518$) and Switzerland ($n = 494$).

To determine the frequency of G191R in non Europeans, we studied 948 individuals of African descent originating from Benin ($n = 161$), Cameroon ($n = 412$), Ethiopia ($n = 155$), and Ecuador ($n = 220$). Detailed characteristics of affected individuals are available online as **Supplementary Table 2** online.

Mutation screening. Oligonucleotide sequences and PCR conditions for *PRSS2* are available online as **Supplementary Methods**. Sequencing of *PRSS1* was performed as described previously[9]. We analyzed DNA by sequencing both strands of the entire coding region.

Melting curve analysis. We performed melting curve analysis to detect the G191R variant using a pair of fluorescence resonance energy transfer (FRET) probes and the LightCycler (Roche Diagnostics). The sensor probe was 5′-CAGGAATCCTTGCCTCCCTC-FL-3′ (FL: 5,6-carboxyfluorescein attached to 3′-*O*-ribose) and the anchor probe was 5′-LC-AGGAAGCCCACACAGAA CATGTTGTTG-ph-3′ (LC: LightCycler Red 640 attached to 5′ terminus; ph: 3′ phosphate). The sensor probe was complementary to the wild-type sequence resulting in an allele-specific melting curve (62.5 °C versus 56 °C). For further details on melting curve analysis and assay validation see the **Supplementary Methods** available online.

Haplotype analysis. For haplotype analysis we selected a total of 17 SNPs (going from *rs2156965*, located 117 kb centromeric from *PRSS2*, to *rs985180*, located 63 kb telomeric from *PRSS2*; **Fig. 1**) from HapMap. We determined the genotypes of 95 subjects from 36 families with at least a carrier of G191R by direct DNA sequencing. In five cases, we were unable to determine the haplotype. Oligonucleotide sequences and PCR conditions for amplifying the five SNPs defining Block 1 (*rs10273639*, *rs1985888*, *rs2855983*, *rs997222*, *rs2367487*) as well as the two nearest SNPs flanking Block 1 (*rs1964986* and *rs2071361*) are available online in **Supplementary Methods**. Pairwise LD was calculated and visualized using the Haploview program[10].

Mutation modeling. We modeled the effect of the G191R variant using the human cationic trypsin crystal structure[11]. The image was rendered using DeepView/Swiss-PdbViewer (v. 3.7)[12].

Generation of recombinant PRSS2 and activation assays. Site-directed mutagenesis, recombinant expression, *in vitro* refolding and purification of human trypsinogens were carried out as described previously[13,14]. Recombinant trypsinogen preparations (2 μM final concentration) were incubated at 37 °C, in 0.1 M Tris-HCl (pH 8.0) and 1 mM $CaCl_2$, in a final volume of 100 μl. Activation was initiated with 20 ng/ml bovine enteropeptidase (Biozyme) or 0.1 μM trypsin (final concentrations). Aliquots of 2.5 μl were withdrawn from reaction mixtures at indicated times, and trypsin activity was determined with 0.14 mM (final concentration) *N*-CBZ-Gly-Pro-Arg-*p*-nitroanilide. Activity was expressed as percentage of the potential total activity, which was determined on a separate wild-type trypsinogen sample after activation with enteropeptidase at 22 °C in 0.1 M Tris-HCl (pH 8.0) and 10 mM Ca^{2+}. For electrophoretic gel analysis, samples were precipitated with 10% trichloroacetic acid (final concentration), resolved by 12% SDS-PAGE under reducing conditions and stained with Coomassie blue, as described earlier[15].

Mass spectrometric determination of cleavage sites. Identification of the cleavage sites within Arg191 trypsinogen was performed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Wild-type and Arg191 trypsinogens were incubated in 10 μl of 50 mM ammonium bicarbonate buffer, pH 8.0, containing 10 mM $CaCl_2$ and 10 milliunits enteropeptidase (Sigma). At the indicated time points, 0.5-μl aliquots were taken and subsequently cocrystallized with 0.5 μl sinapinic acid (saturated in 70% acetonitrile) on a SCOUT 384 -MALDI-Target. Mass spectrometry was performed on a MALDI-TOF-MS (Reflex III, Bruker Daltonics) instrument in linear mode with internal calibration. Annotation of trypsinogen fragments was performed with BioTools 2.0 software (Bruker Daltonics).

Statistics. The significance of the differences between mutation frequencies in affected individuals and controls as well as between different control populations were tested by two-tailed Fisher's Exact tests, and additional odds ratios were calculated using SAS/STAT software (v 9.1).

Accession codes. Entrez nucleotide: protease, serine 2 (*PRSS2*): U66061. Protein Data Bank: trypsin crystal structure, 1TRN.

URLs. HapMap: http://www.hapmap.org. DeepView/Swiss-PdbViewer v. 3.7: http://www.expasy.org/spdbv/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the many patients who have participated in this study. We are especially grateful to V. Kielstein (Magdeburg, Germany) and L. Klomp (University Medical Center Utrecht, Utrecht, The Netherlands) for providing

blood samples from control subjects. The initial experiments were supported by the DFG (Wi 2036/1-1). This work was supported by the Sonnenfeld-Stiftung, Berlin, Germany (to H.W.), the US National Institutes of Health (NIH) (grant DK058088 to M.S.-T.), INSERM (Institut National de la Santé et de la Recherche Médicale) and the Programme Hospitalier de Recherche Clinique (grant PHRC R 08-04 to C.F.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Whitcomb, D.C. *et al.* Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat. Genet.* **14**, 141–145 (1996).
2. Witt, H. *et al.* Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat. Genet.* **25**, 213–216 (2000).
3. Scheele, G., Bartelt, D. & Bieger, W. Characterization of human exocrine pancreatic proteins by two-dimensional isoelectric focusing/SDS gel electrophoresis. *Gastroenterology* **80**, 461–473 (1981).
4. Chen, J.M. & Férec, C. Human trypsins. in *Handbook of Proteolytic Enzymes* 2nd edn. (eds. Barrett, A.J. *et al.*) 1489–1493 (Elsevier, London, 2004).
5. Colomb, E. & Figarella, C. Comparative studies on the mechanism of activation of the two human trypsinogens. *Biochim. Biophys. Acta* **571**, 343–351 (1979).
6. Colomb, E., Guy, O., Deprez, P., Michel, R. & Figarella, C. The two human trypsinogens: catalytic properties of the corresponding trypsins. *Biochim. Biophys. Acta* **525**, 186–193 (1978).
7. Rowen, L., Koop, B.F. & Hood, L. The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science* **272**, 1755–1762 (1996).
8. Kénesi, E., Katona, G. & Szilágyi, L. Structural and evolutionary consequences of unpaired cysteines in trypsinogen. *Biochem. Biophys. Res. Commun.* **309**, 749–754 (2003).
9. Witt, H., Luck, W. & Becker, M. A signal peptide cleavage site mutation in the cationic trypsinogen gene is strongly associated with chronic pancreatitis. *Gastroenterology* **117**, 7–10 (1999).
10. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
11. Gaboriaud, C., Serre, L., Guy-Crotte, O., Forest, E. & Fontecilla-Camps, J.-C. Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151. *J. Mol. Biol.* **259**, 995–1010 (1996).
12. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
13. Sahin-Tóth, M. Human cationic trypsinogen. Role of Asn-21 in zymogen activation and implications in hereditary pancreatitis. *J. Biol. Chem.* **275**, 22750–22755 (2000).
14. Sahin-Tóth, M. & Tóth, M. Gain-of-function mutations associated with hereditary pancreatitis enhance autoactivation of human cationic trypsinogen. *Biochem. Biophys. Res. Commun.* **278**, 286–289 (2000).
15. Kukor, Z., Tóth, M. & Sahin-Tóth, M. Human anionic trypsinogen. Properties of autocatalytic activation and degradation and implications in pancreatic diseases. *Eur. J. Biochem.* **270**, 2047–2058 (2003).