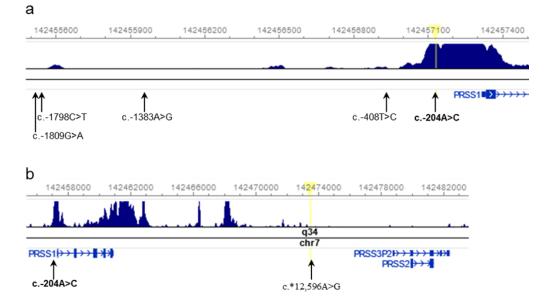
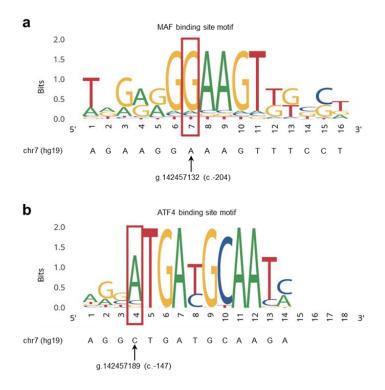
Location	Variant				Clinical genetic data	gpAF (hspAF) ^a	Classification (for CP)
	c.nomenclature (NM_002769.5)	g.nomenclature (chr7, hg19)	Amino acid change	rs number			
Exon 2	c.111C>A	g.142458476C>A	p.Y37*	Not available	Found in a subject with alcoholism but without pancreatitis [1]	Not available	Benign (protective)
Exon 2	c.166C>T	g.142458531C>T	p.Q56*	rs147366981	Identified in a Chinese patient with pancreatic cancer [2]	Not available	Benign (protective)
Exon 2	c.171G>A	g.142458536G>A	p.W57*	Not available	Identified in a Chinese patient with pancreatic ductal adenocarcinoma [3]	Not available	Benign (protective)
Exon 3	c.204C>T	g.142459628C>T	p.R68=	Not available	Identified in a Caucasian patient with idiopathic pancreatitis ^b [4]	Not available	Benign
Exon 3	c.256C>T	g.142459680C>T	p.Q86*	rs146398318	Identified in a English index patient with hereditary diffuse gastric cancer [5] and in a patient with pancreatic ductal adenocarcinoma [6]	0.00005303 (0.0001006, European (non-Finnish)	Benign (protective)
Exon 3	c.273C>A ^c	g.142459697C>A	p.A91=	Not available	In a three-generation Chinese family with autoimmune pancreatitis, three patients and three healthy members were available for genetic analysis. Only the index patient (III-2) carried this variant [7]	Not available	Benign
Exon 3	c.417C>T	g.142459841C>T	p.C139=	rs141847266	Identified in a healthy Chinese subject [8]	0.0003820 (0.004812, East Asian)	Benign
Exon 4	c.480C>A	g.142460307C>A	p.C160*	Not available	Identified in a patient with pancreatic adenocarcinoma [6]	Not available	Benign (protective)
Exon 4	c.486T>C (c.486C>T) ^d	g.142460313T>C	p.D162=	rs6666 ($R^2 = 0.9743$, relative to c408T>C)	Described in multiple studies. In the first report, the variant was identified in some members of a family with HCP caused by p.Asn29Ile [9]	0.3878 (0.4301, non- Finnish European)	Benign
Exon 4	c.491delC	g.142460318del	p.P164Lfs*3	rs775936808	Identified in a patient with prostate cancer [10]	0.00001414 (0.00004004, African/African American)	Benign (protective)
Exon 4	c.567T>C	g.142460394T>C	p.L189=	rs377124851	Identified in a subject with aplastic anemia [11]	Not available	Benign
Exon 5	c.738T>C (c.738C>T) ^d	g.142460865T>C	p.N246=	rs6667 ($R^2 = 0.9858$, relative to c408T>C)	Described in multiple studies [12]. In the first report, the variant was found in some members of two families with HCP caused by <i>PRSS1</i> p.Asn29IIe [13]	0.5097 (0.6048, European (Finnish))	Benign

Supplementary Table 1. Classification of PRSS1 nonsense, frameshift deletion and silent variants

^aIn accordance with gnomAD v2.1.1 [14]. ^bThierry Bienvenu, personal communication. ^cErroneously described as c.279C>A in the original report [7]. ^dAlternate name in accordance with NG_001333.2 (Homo sapiens T cell receptor beta locus on chromosome 7).



Supplementary Figure 1. Correlation of the six common 5' and 3' variants with DNase I-accessible DNA regions in the *PRSS1* locus. (a) Locations of the five 5' variants. (b) Location of the single 3' variant (c.-204A>C is included for the sake of comparison). Nucleotide coordinates refer to chromosome 7 (hg19). The experimentally demonstrated functional c.-204A>C is highlighted in bold. The Figure was obtained from the portal site of the NIH Roadmap Epigenomics Mapping Consortium website. DNase I-accessible DNA regions were determined in the pancreatic tissue of a donor by means of DNase-seq.



Supplementary Figure 2. Transcription factor binding sites affected by two experimentally demonstrated functional *PRSS1* 5' variants. (a) c.-204A>C. (b) c.-147C>T. In each panel, the upper part shows the transcription factor binding site motif annotated by RegulomDB whilst the lower portion shows the corresponding chr7 genomic sequences. Variant-affected nucleotide positions are indicated by arrows.

а						c.40+1G>A ↓			
1	aggcacactc	taccaccatg	aatccactcc	tgatccttac	ctttgtggca	gctgctc <mark>g</mark> tq			
61	<u>a</u> GTatcatgc	cctgcctcag	gccccaacca	ccccccgtt	cctggccgac	aaatgccctt			
121	ccattcttac	cacctctcct	cttttgactg	tgctctgata	ttccgtttcc	tccatctggc			
181	atatctcctt	cccatcctcc	ttgggctctt	tttaagtctc	acctgttcac	cttctccttg			
241	acttcactcc	caccactgtc	attcatccat	atccgagttg	tggttggaga	agctgggaag			
301	ggggccaggt	ggggctgtcc	cacgaaatga	agcagcaggc	ttcaggcttg	gctccgacag			
361	caacagaata	gcaccactat	agctgctcct	aacctcgaat	gcacctgggg	aggttgaaaa			
421	attactcatg	ccagagactc	agctctagaa	attctaactt	caattgtctg	gggtagagct			
481	tgtgttctgg	gcttttaagc	ttcccaggtg	atttttaata	tttccagcca	tgcagccaag			
541	gttaagaatt	gctgtcctat	tggccaataa	caaagtctac	ctttgttctg	ccaaagtgag			
601	cctgggggct	gccctcaact	ctgccctgac	tgcacagatc	tgagctatgg	gggaaggtgg			
661	tcatggccag	gtctatgcag	acagggggtt	ttcctagctt	ggcaaaagaa	tcctgacaat			
721	ccagggccca	aatagccagg	ggaagtacac	aggtgatgaa	taaaagagag	aagcactcag			
781	tgggagagac	aaccacatcc	caactcctat	cccactggaa	gcattgtgag	gacattcctt			
841	gcgacttcag	cctggtgacc	ccaggagagc	tcggatcctc	cgcagggtac	ctagctatgt			
901	gccctgcagg	cacagagact	tgggagccac	aggcagtgat	gatcaccagg	ggtggcagag			
961	ctccctccct	tgcctagcct	cactgtgctt	gttaaggatt	tctaattagc	agaaagcaat			
1021	cacaggctgg	gagcgccacc	cctaacatgc	tattgacttg	ccttctccct	tcccatctcc			
1081	actccag ttg	ctgcccctt	<mark>tga</mark> tgatgat	gacaagatcg	ttgggggcta	caactgtgag			
1141	gagaattctg	tcccctacca	ggtgtccctg	aattctggct	accacttctg	tggtggctcc			
1201	ctcatcaacg	aacagtgggt	ggtatcagca	ggccactgct	acaagtc gta	agtgtggggc			
1261	ccccgactgc	aaagctcccg	gccagtctgc	ctgggagagc	ttggcttcag	cccagggaac			
b atgaatccactcctgatccttacctttgtggcagctgctc <u>atqa</u> ttgctgccccctttgatgatga M N P L L I L T F V A A A H D C C P L *									

Supplementary Figure 3. Illustration of the predicted functional consequence of the c.40+1G>A variant. (a) Sequence shown is the partial genomic sequence of the *PRSS1* gene (NG_001333.2), with exons 1 and 2 being highlighted in bold. The translation initiation codon ATG is highlighted in blue. c.40+1G>A was predicted by SpliceAI to disrupt the splice donor site of intron 1 whilst activating a downstream cryptic splice donor site (highlighted in capital letters and red). Consequently, the mutant transcript would be predicted to include the first four nucleotides (underlined) of intron 1, thereby creating a premature translation termination codon (highlighted in red) within exon 2. (b) Protein product encoded by the predicted mutant transcript. The retained four nucleotides are underlined, with the variant being highlighted in red.

```
1081 actccagttg ctgccccctt tgatgatgat gacaagatcg ttggggggcta caactgtgag
 1141 gagaattetg teeectacea ggtgteeetg aattetgget accaettetg tggtggetee
 1201 ctcatcaacg aacaGTgggt ggtatcagca ggccactgct acaagtcgta agtgtggggc
                                                        c.200+1G>A
1261 ccccqactqc aaagctcccq qccaqtctqc ctqqqaqaqc ttqqcttcaq cccaqqqaac
b
    MNPLLILTFVAAALAAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGSLINEQWVVSAGHCYKSRIQVRLGEHNIEV
WΤ
                                                                         80
Mut MNPLLILTFVAAALAAPFDDDDKIVGGYNCEENSVPYOVSLNSGYHFCGGSLINEH-----RIOVRLGEHNIEV
                                                                         69
                                          *********
WΤ
    LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELQC 160
Mut LEGNEOFINAAKIIRHPOYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELOC 149
WΤ
    LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK 240
Mut LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK 229
     247
WΤ
    NTIAANS
Mut NTIAANS
                                                                        236
```

а

Supplementary Figure 4. Illustration of the predicted functional consequence of the c.200+1G>A variant. (a) Sequence shown is the partial genomic sequence of the *PRSS1* gene (NG_001333.2), with exon 2 being highlighted in bold. c.200+1G>A was predicted by SpliceAI to disrupt the wild-type splice donor site of intron 2 whilst activating an upstream cryptic splice donor site within exon 2 (highlighted in red capital letters). This would be predicted to lead to a deletion of 33 bp (barred) in the mutant transcript. (b) Alignment of the wild-type and predicted mutant PRSS1 amino acid sequences. The catalytic triad residues, His^{63} , Asp^{107} and Ser^{200} , are highlighted in bold in the context of the wild-type sequence. Loss of His^{63} would render the mutant enzyme inactive.

а 1 aggcacacte taccaccatg aatecactee tgateettae etttgtggca getgetegtg 61 agtatcatge cetgeeteag geeceaacea ceeceegtt cetggeegae xxxxxxxxx 1021 cacaggetgg gagegeeace estaacatge tattgaettg estteteest 1081 actccagttg otgococott tgatgatgat gacaagateg ttggggggota caactgtgag 1141 gagaattotg tocootacca ggtgtocotg aattotggot accasttotg tggtggotco 1201 otoatoaacg aacagtgggt ggtateAGca ggccactgct acaagtcgta agtgtggggc 1261 ccccgactgc aaagctcccg gccagtctgc ctgggagagc ttggcttcag cccagggaac 1321 tactgaggtt gggtaagatg gatgggagag gtggtggaga agaaaacttg b 1 acactetace accatgaate casteetgat cettacettt gtggeagetg ete $\frac{1}{2}$ 61 cccctttgat gatgatgaca agatcgttgg gggctacaac tgtgaggaga attctgtccc 121 ctaccaggtg tccctgaatt ctggctacca cttctgtggt ggctccctca tcaacgaaca 181 gtgggtggta tcagcaggcc actgctacaa gtcccgcatc caggtgagac tgggagagca 241 caacatcgaa gtcctggagg ggaatgagca gttcatcaat gcagccaaga tcatccgcca 301 cccccaatac gacaggaaga ctctgaacaa tgacatcatg ttaatcaagc tctcctcacg 361 tgcagtaatc aacgcccgcg tgtccaccat ctctctgccc accgcccctc cagccactgg 421 cacgaagtgc ctcatctctg gctggggcaa cactgcgagc tctggcgccg actacccaga 481 cgagetgeag tgeetggatg eteetgtget gageeagget aagtgtgaag eeteetaece 541 tggaaagatt accagcaaca tgttctgtgt gggcttcctt gagggaggca aggattcatg 601 tcagggtgat tctggtggcc ctgtggtctg caatggacag ctccaaggag ttgtctcctg 661 gggtgatggc tgtgcccaga agaacaagcc tggagtctac accaaggtct acaactatgt 721 gaaatggatt aagaacacca tagctgccaa tagctaaagc ccccagtatc tcttcagtct 781 ctataccaat aaagtgaccc tgttctcac С Pretrypsinogen Trypsinogen -> Trypsin MVPLL1LTFVAAALAAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGSLINEQWVVSAGHCYKSRIQVRLGEHNIEV WТ 80 Mut MNPLLILTFVAAA-----PGHCYKSRIOVRLGEHNIEV 33 -*** LEGNEQFINAAKIIRHPQYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELQC 160 WΤ LEGNEOFINAAKIIRHPOYDRKTLNNDIMLIKLSSRAVINARVSTISLPTAPPATGTKCLISGWGNTASSGADYPDELOC 113 Mut WT LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK 240 Mut LDAPVLSQAKCEASYPGKITSNMFCVGFLEGGKDSCQGDSGGPVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK 193 WT NTIAANS 247 NTIAANS 200 Mut

Supplementary Figure 5. Illustration of the functional effect of the *PRSS1* c.200+64_68delCCCAG variant. (a) Sequence shown is partial genomic sequence of the *PRSS1* gene (NG_001333.2), with exons 1 and 2 being highlighted in bold. The translation initiation codon ATG is highlighted in blue. The pentanucleotide deletion CCCAG at the DNA level is highlighted in red and underlined. This deletion, for unknown reasons, inactivated the physiological intron 1 acceptor splice site and activated a cryptic acceptor splice site within exon 2 (highlighted in bold and red). The aberrantly spliced (and predicted to be excluded) sequence within exon 2 is barred. (b) Sequence shown is the mRNA sequence of the *PRSS1* gene (NM_002769.5). The aforementioned aberrantly spliced sequence within exon 2 (in-frame deletion of 141 nucleotides) is barred. The translation initiation codon (for both wild-type and mutant transcripts) are highlighted in blue. (c) Alignment of the wild-type and predicted mutant PRSS1 amino acid sequences. The catalytic triad residues, His⁶³, Asp¹⁰⁷ and Ser²⁰⁰, are highlighted in red in the context of the wild-type sequence.

Supplementary references

- 1. Chen JM, Le Maréchal C, Lucas D, Raguénès O, Férec C. "Loss of function" mutations in the cationic trypsinogen gene (*PRSS1*) may act as a protective factor against pancreatitis. Mol Genet Metab. 2003;79(1):67-70.
- 2. Liu Q, Guo L, Zhang S, Wang J, Lin X, Gao F. *PRSS1* mutation: a possible pathomechanism of pancreatic carcinogenesis and pancreatic cancer. Mol Med. 2019;25(1):44.
- 3. Yin L, Wei J, Lu Z, Huang S, Gao H, Chen J, et al. Prevalence of germline sequence variations among patients with pancreatic cancer in China. JAMA Netw Open. 2022;5(2):e2148721.
- 4. Girodon E, Rebours V, Chen JM, Pagin A, Levy P, Férec C, et al. Clinical interpretation of *PRSS1* variants in patients with pancreatitis. Clin Res Hepatol Gastroenterol. 2021;45(1):101497.
- 5. Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, et al. Hereditary diffuse gastric cancer syndrome: *CDH1* mutations and beyond. JAMA Oncol. 2015;1(1):23-32.
- Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol. 2017;35(30):3382-90.
- 7. Gao F, Li YM, Hong GL, Xu ZF, Liu QC, He QL, et al. *PRSS1_p.Leu81Met mutation results in autoimmune pancreatitis*. World J Gastroenterol. 2013;19(21):3332-8.
- 8. Gou S, Yu J, Wang C, Liu T, Cui P, Li X. Three female familial cases of solid pseudopapillary tumors with a protease serine 1 gene mutation. Pancreas. 2013;42(1):168-73.
- Gorry MC, Gabbaizedeh D, Furey W, Gates LK, Jr., Preston RA, Aston CE, et al. Mutations in the cationic trypsinogen gene are associated with recurrent acute and chronic pancreatitis. Gastroenterology. 1997;113(4):1063-8.
- 10. Mijuskovic M, Saunders EJ, Leongamornlert DA, Wakerell S, Whitmore I, Dadaev T, et al. Rare germline variants in DNA repair genes and the angiogenesis pathway predispose prostate cancer patients to develop metastatic disease. Br J Cancer. 2018;119(1):96-104.
- 11. Zhang Y, Zhang Y, Ge H, Li N, Liu C, Wang T, et al. Identification of potential pathogenic genes for severe aplastic anemia by whole-exome sequencing. J Clin Lab Anal. 2022;36(5):e24438.
- 12. Database of Genetic Risk Factors in Chronic Pancreatitis. https://pancreasgenetics.org/. Last accessed 29 March 2023.
- 13. Teich N, Mossner J, Keim V. Mutations of the cationic trypsinogen in hereditary pancreatitis. Hum Mutat. 1998;12(1):39-43.
- 14. gnomAD (Genome Aggregation Database). https://gnomad.broadinstitute.org/. Accessed 16 March 2023.