OPEN ACCESS

ORIGINAL ARTICLE

Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis

Jonas Rosendahl,[1,2] Holger Kirsten,[3,4,5] Eszter Hegyi,[6] Peter Kovacs,[7] Frank Ulrich Weiss,[8] Helmut Laumen,[9] Peter Lichtner,[10] Claudia Ruffert,[1] Jian-Min Chen,[11] Emmanuelle Masson,[11] Sebastian Beer,[2] Constantin Zimmer,[2] Katharina Seltsam,[2] Hana Algül,[12] Florence Bühler,[9] Marco J Bruno,[13] Peter Bugert,[14] Ralph Burkhardt,[4,15] Giulia Martina Cavestro,[16] Halina Cichoz-Lach,[17] Antoni Farré,[18] Josef Frank,[19] Giovanni Gambaro,[20] Sebastian Gimpfl,[9] Harald Grallert,[21,22,23] Heidi Griesmann,[1] Robert Grützmann,[24] Claus Hellerbrand,[25] Péter Hegyi,[26,27] Marcus Hollenbach,[1] Sevastitia Iordache,[28] Grazyna Jurkowska,[29] Volker Keim,[2] Falk Kiefer,[30] Sebastian Krug,[1] Olfert Landt,[31] Milena Di Leo,[16] Markus M Lerch,[8] Philippe Lévy,[32] Markus Löffler,[3,4] Matthias Löhr,[33] Maren Ludwig,[9] Milan Macek,[34] Nuria Malats,[35,36] Ewa Malecka-Panas,[37] Giovanni Malerba,[38] Karl Mann,[30] Julia Mayerle,[39] Sonja Mohr,[9] Rene H M te Morsche,[40] Marie Motyka,[9] Sebastian Mueller,[41] Thomas Müller,[42] Markus M Nöthen,[43,44] Sergio Pedrazzoli,[45] Stephen P Pereira,[46] Annette Peters,[22,23,47] Roland Pfützer,[48] Francisco X Real,[36,49,50] Vinciane Rebours,[32] Monika Ridinger,[51] Marcella Rietschel,[19] Eva Rösmann,[9] Adrian Saftoiu,[28] Alexander Schneider,[52] Hans-Ulrich Schulz,[53] Nicole Soranzo,[54,55] Michael Soyka,[56] Peter Simon,[8] James Skipworth,[57] Felix Stickel,[58] Konstantin Strauch,[59,60] Michael Stumvoll,[7,61] Pier Alberto Testoni,[16] Anke Tönjes,[61] Lena Werner,[9] Jens Werner,[62] Norbert Wodarz,[51] Martin Ziegler,[9] Atsushi Masamune,[63] Joachim Mössner,[2] Claude Férec,[11] Patrick Michl,[1] Joost P H Drenth,[40] Heiko Witt,[9] Markus Scholz,[3,4] Miklós Sahin-Tóth,[6] On behalf of all members of the PanEuropean Working group on ACP

Correspondence to
Professor Jonas Rosendahl, Department of Internal Medicine, Neurology and Dermatology Division of Gastroenterology and Rheumatology, University Clinic of Leipzig Liebigstraße 20, 04103 Leipzig, Germany; jonas.rosendahl@medizin.uni-leipzig.de

JR and HK contributed equally.

JPHD, HW, MS and MS-T contributed equally to the supervision of the work.

Check for updates

ABSTRACT

Objective Alcohol-related pancreatitis is associated with a disproportionately large number of hospitalisations among GI disorders. Despite its clinical importance, genetic susceptibility to alcoholic chronic pancreatitis (CP) is poorly characterised. To identify risk genes for alcoholic CP and to evaluate their relevance in non-alcoholic CP, we performed a genome-wide association study and functional characterisation of a new pancreatitis locus.

Design 1959 European alcoholic CP patients and population-based controls from the KORA, LIFE and INCIPE studies (n=4708) as well as chronic alcoholics from the GESGA consortium (n=1332) were screened with Illumina technology. For replication, three European cohorts comprising 1650 patients with non-alcoholic CP and 6695 controls originating from the same countries were used.

Results We replicated previously reported risk loci *CLDN2-MORC4*, *CTRC*, *PRSS1-PRSS2* and *SPINK1* in alcoholic CP patients. We identified *CTRB1-CTRB2* (chymotrypsin B1 and B2) as a new risk locus with lead single-nucleotide polymorphism (SNP) *rs8055167* (OR

1.35, 95% CI 1.23 to 1.6). We found that a 16.6 kb inversion in the *CTRB1-CTRB2* locus was in linkage disequilibrium with the CP-associated SNPs and was best tagged by *rs8048956*. The association was replicated in three independent European non-alcoholic CP cohorts of 1650 patients and 6695 controls (OR 1.62, 95% CI 1.42 to 1.86). The inversion changes the expression ratio of the CTRB1 and CTRB2 isoforms and thereby affects protective trypsinogen degradation and ultimately pancreatitis risk.

Conclusion An inversion in the *CTRB1-CTRB2* locus modifies risk for alcoholic and non-alcoholic CP indicating that common pathomechanisms are involved in these inflammatory disorders.

INTRODUCTION

Chronic pancreatitis (CP) is a relapsing, progressive inflammatory disorder of alcoholic (ACP), idiopathic or hereditary aetiology. In about half of non-alcoholic CP (NACP) cases, a genetic background has been identified with mutations

Significance of this study

What is already known on this subject?
► Genetic risk underlying alcoholic chronic pancreatitis (CP) is poorly understood. In contrast, the genetic basis of non-alcoholic CP has been more comprehensively characterised.
► Alcohol abuse is a predominant cause of CP; however, only a small percentage of alcohol abusers develop the disease, suggesting that genetic susceptibility may contribute to pathogenesis.
► A genome-wide association study (GWAS) reported variants in the *PRSS1-PRSS2* and *CLDN2-MORC4* loci associated with CP. More recent studies indicate that these variants may have the strongest effect in alcoholic CP.

What are the new findings?
► This is the largest European GWAS in patients with alcoholic CP. The results replicate the reported associations with variants in the *CLDN2-MORC4*, *CTRC*, *PRSS1-PRSS2* and *SPINK1* loci.
► We identified *CTRB1-CTRB2* (chymotrypsin B1 and B2) as a new risk locus for alcoholic CP and found that a 16.6 kb inversion in the *CTRB1-CTRB2* locus was in linkage disequilibrium with the associated single-nucleotide polymorphisms. The association was replicated in non-alcoholic CP.
► The inversion changes the expression ratio of the CTRB1 and CTRB2 isoforms and thereby affects protective trypsinogen degradation and ultimately pancreatitis risk.

How might it impact on clinical practice in the foreseeable future?
► The results indicate that alcoholic and non-alcoholic CP share common pathomechanisms.
► As the same variants influence development of alcoholic and non-alcoholic CP, therapeutic approaches should be guided by disease mechanism rather than aetiology.
► The identified risk variants explain about 18% of the variance in alcoholic CP and may serve as the basis for risk assessment in the clinical setting.

in risk genes *CTRC* (chymotrypsin C), *PRSS1* (serine protease 1, cationic trypsinogen) and *SPINK1* (serine protease inhibitor Kazal type 1).[1] Most changes in these genes increase ectopic intra-pancreatic activation of the digestive serine protease trypsin.[2–5] However, the relevance of such a trypsin-dependent mechanism in ACP has remained contentious as association with high-effect *CTRC*, *PRSS1* and *SPINK1* variants is absent or considerably weaker than in NACP.[5 6] Similarly, rare *PRSS1* and *CPA1* (carboxypeptidase A1) variants that result in misfolding and endoplasmic reticulum stress are associated with NACP but not with ACP.[7 8] The so far unidentified genetic susceptibility in ACP is an intriguing observation as only a small percentage of alcohol abusers develop ACP, suggesting that factors other than alcohol may contribute to disease onset.[9 10]

Indeed, a recent genome-wide association study (GWAS) revealed association of common variants in the *CLDN2-MORC4* and the *PRSS1-PRSS2* loci with ACP and NACP.[11] These findings were first replicated in a large European cohort where association was strongest in the ACP group followed by similar observations in Japanese and Indian CP cohorts.[12–15] Functional studies indicated that the protective *PRSS1* promoter variant reduces transcription that should result in lower intra-pancreatic trypsinogen levels.[16]

Here, we present the largest European multicentre GWAS combining 1959 ACP cases from nine countries using Illumina chips. As controls, chronic alcoholics without CP (n=1332) and population-based controls (n=4708) with data on alcohol consumption were used (total n=6040).

METHODS
Study population
In all participating study centres, the corresponding medical ethical review committees approved the study. CP was defined by a typical clinical course with recurrent attacks or chronic pain and characteristic morphological changes in imaging studies as well as functional impairment with exocrine and/or endocrine insufficiency. CP was considered as ACP if alcohol consumption was >60 g for females and >80 g for males per day over at least 2 years. The cohorts are summarised in table 1 and in online supplementary table S1.

Table 1 Description of the cohorts included in the analysis				
Cohort type	**No.**	**Alcohol abuse**	**Age (range)**	**Male sex (%)**
Initial GWAS cohort				
ACP cases (Pan-European Working Group)	1959	Yes	50 (17–98)	1674 (85.5%)
Controls (KORA, INCIPE)	2637	–	49 (25–84)	1073 (62.0%)
Controls (GESGA and KORA)	1488	Yes	48 (26–74)	1474 (99.1%)
Controls (KORA)	1915	No	47 (25–74)	686 (35.8%)
Replication cohort 1				
NACP cases (Pan-European Working Group)	584	No	21 (1–71)	287 (49.1%)
Controls (LIFE)	4892	–	63 (19–82)	2379 (48.6%)
Replication cohort 2				
NACP cases (France, Brest)	546	No	18 (1–72)	281 (51.5%)
Controls (France, Brest)	1043*	–	30 (17–75)	295 (50.6%)
Replication cohort 3				
NACP cases (Germany, Greifswald)	520	No	41 (1–86)	288 (55.4%)
Controls (Germany, Greifswald)	760	–	32 (18–68)	465 (61.0%)

Further details of cohorts are provided in online supplementary table S1.
*In the French control cohort, age and gender data were available for 583 subjects only.
ACP, alcoholic chronic pancreatitis; GWAS, genome-wide association study; No., number of individuals after quality control filtering; NACP, non-alcohol-related chronic pancreatitis.

Genotyping, quality control and GWAS statistics

Genotyping was performed on Illumina BeadChip arrays (Illumina, San Diego, California, USA). The ACP samples of the screening cohort and NACP samples of replication cohort 1 were genotyped at the Helmholtz Center Munich (Dr P. Lichtner). The study design is summarised in online supplementary figure S1.

All analyses were performed using PLINK V.1.9[17] and R (www.R-project.org). Data were filtered to achieve an individual-wise and single-nucleotide polymorphism (SNP)-wise call rate >0.99.[18] Genotyped sex was determined in each sample from gonosomal data. Related individuals were excluded (pi-hat >0.185). Heterozygosity outliers defined by heterozygosity more extreme than median ±3 IQRs and ethnic outliers according to Price et al[18] with 6 SD criterion in principal component analysis were not used for further analysis. Hence, the 1959 ACP samples that passed quality control as well as controls (n=6040) derived from different consortia (KORA S3/F3 Illumina Omni data set from Augsburg, Germany; GESGA consortia data set from Mannheim, Germany; INCIPE Illumina Omni data set from Verona, Italy) were included for further analyses.

Imputation was performed with SHAPEIT V.2 and IMPUTE V.2.3.0 applying the 1000Genomes reference phase 1, V. 3[19] and logistic regression was applied with the first three principal components of the SNP data included as covariates to account for possible population stratification resulting in a standardised overall inflation factor of 1.02.[20]

Imputation of the 16.6 kb inversion for further analysis

Imputation of a 16.6 kb inversion[21] in the *CTRB1-CTRB2* locus within the screening cohort was performed using a 5 Mb region on chromosome 16 between 71 310 697 bp and 76 045 524 bp including 475 SNPs overlapping between all cohorts and the reference. As reference panel, quality filtered data of HumanOmniExpress BeadChips from 227 CP patients with successfully genotyped inversion genotypes were used (for details, see online supplementary file 1). To improve accuracy, the IMPUTE2 parameters number of Markov chain Monte Carlo iterations were set to 30 and number of hidden Markov Model states were increased to 200.

Calculation of variance of ACP

Joint variance explained by all identified loci was measured by calculating McFadden's pseudo-R^2. Here, we compared a logistic regression including the three principal components and literature SNPs rs497078, rs17107315, rs10273639, rs12688220 and the inversion as independent variables with a logistic regression model including the first three principal components only.

Replication genotyping

Details of polymerase chain reactions (PCR) and melting curve assays used for analysis of the *CTRB1-CTRB2* locus can be found in the online supplementary file 1. In the first replication cohort of the German NACP patients and controls, chip data were used with similar preprocessing as described (n=584 NACP patients; n=4892 population-based controls from LIFE, the latter were genotyped using Affymetrix AXIOM-CEU (Affymetrix, St. Clara, California, USA) genome-wide SNP array[22]). Thereby, imputation on 1000Genomes reference phase 1, V. 3 was performed for a 5 Mb region on chromosome 16 between 71 310 697 bp and 76 045 524 bp using 285 SNPs fulfilling quality control in all individuals.

Further replication in two independent NACP cohorts from Germany (Greifswald cohort, 520 patients, 760 controls)

and France (Brest cohort, 546 patients, 1043 controls) was carried out by genotyping SNPs with the melting curve assay (see online supplementary table S2).

Measurement of *CTRB1* and *CTRB2* mRNA expression

Samples of human pancreatic cDNA prepared from pancreatic exocrine fractions discarded after islet isolation were kind gifts from Dr Sohail Husain (Children's Hospital of Pittsburgh) and Dr Rajinder Dawra (University of Minnesota) or were prepared from discarded surgical specimens at the University of Szeged, Hungary. Details of the methods used are summarised in online supplementary file 1.

Trypsinogen activation and degradation

Human cationic trypsinogen (PRSS1) and anionic trypsinogen (PRSS2) were produced in *Escherichia coli*. His-tagged forms of CTRB1 and CTRB2 were expressed in HEK 293 T cells and purified as described previously.[23 24] The effect of CTRB1 and CTRB2 on trypsinogen was determined in autoactivation experiments and degradation assays.

RESULTS

Association of ACP with loci *CLDN2-MORC4, CTRC, PRSS1-PRSS2* and *SPINK1*

We robustly confirmed the association of loci *CLDN2-MORC4* and *PRSS1-PRSS2* reported in a previous GWAS.[11] Our lead SNPs at the *CLDN2-MORC4* (rs12688091) and the *PRSS1-PRSS2* locus (rs2855983) were in linkage disequilibrium (LD) with the previously identified lead SNPs in these loci (see online supplementary table S3). Furthermore, strong association was found with rs545634 in *CTRC* and with rs146437551 in *SPINK1*. Both susceptibility genes were previously identified in candidate gene studies of CP.[3 5 6 25–29] In line with the literature, our top hits at these loci were in strong LD with the most frequently reported SNPs; *CTRC*, rs497078 (c.180C>T, p.G60=) and *SPINK1*, rs17107315 (c.101A>G, p.N34S) (see online supplementary table S3). Association of the lead SNPs at the four risk loci remained essentially unchanged when ACP patients were separately compared with chronic alcoholics and non-alcoholic controls (see online supplementary figure S2).

Association of ACP with an alcohol-dependence locus

We also observed an association when we compared ACP patients versus chronic alcoholics at the known alcohol-dependence locus *ADH1B* (alcohol dehydrogenase 1B) for rs1229984 (OR 2.49, 95% CI 1.84 to 3.39; p=1.8×10^{-8}). This variant (c.143A>G, p.H48R) alters alcohol metabolism and thereby deters from drinking, resulting in the observed lower frequency among alcoholics (see online supplementary figure S3).[30]

Novel association of ACP with the *CTRB1-CTRB2* locus

We discovered a novel association signal at the *CTRB1-CTRB2* (chymotrypsin B1 and B2) locus (figure 1 and table 2 for all top hits) with lead SNP rs8055167 located in intron 1 of *CTRB1* (OR 1.35, 95% CI 1.23 to 1.6; p=4.2×10^{-9}). The association was also observed when ACP patients were compared with alcoholics or population-based controls (see online supplementary figure S2). Regional association plots for the top hits and all associations with p<10^{-5} are summarised in online supplementary figure S4 and online supplementary table S4 and risk estimates for carriers of multiple risk alleles are shown in online supplementary figure S5. Here, the identified risk variants explain about 18% of the variance in ACP.
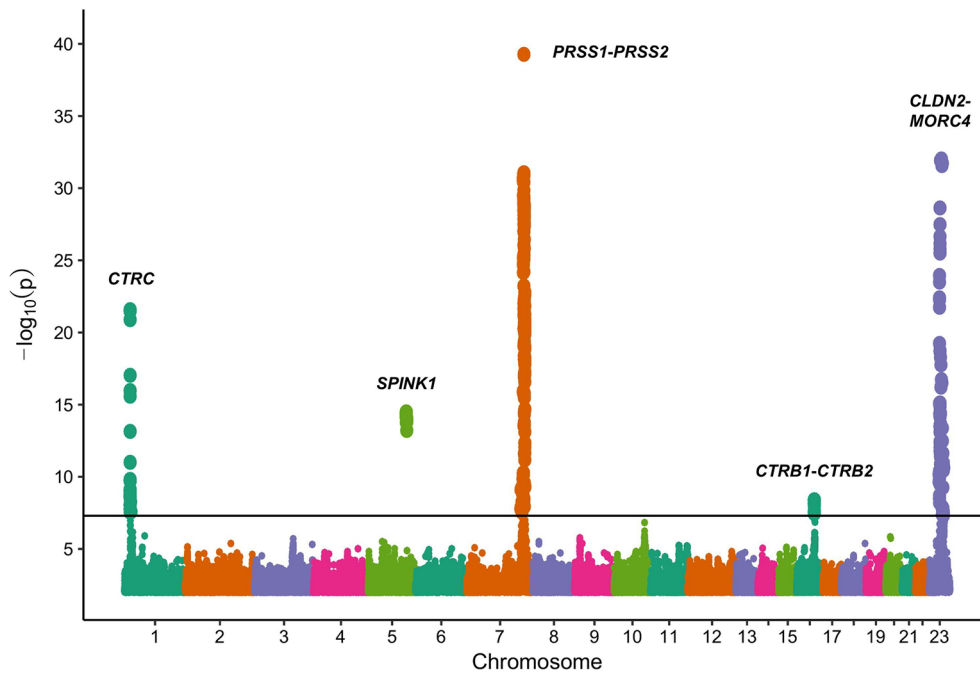
Figure 1 Genome-wide association analysis of 1959 cases with alcoholic chronic pancreatitis and 6040 controls derived from population studies and a cohort of alcohol-dependent patients. Genome-wide significance-level threshold (p=5×10^{-8}) is represented by the black line. Only single-nucleotide polymorphisms that passed quality control are depicted.

Identification of a complex genetic rearrangement in the *CTRB1-CTRB2* locus associated with CP

The *CTRB1-CTRB2* locus harbours complex genomic rearrangement variants which include a 16.6 kb inversion (see figure 2) and a 584 bp deletion in *CTRB2* (not shown).[21] First, we excluded the deletion in *CTRB2* as an explanation for the association as genotyping of 289 ACP patients revealed only negligible LD (R^2=0.12) with our lead SNP *rs8055167*. Subsequently, in order to examine whether our lead SNP is in LD with the 16.6 kb inversion in the *CTRB1-CTRB2* locus, we genotyped 227 ACP patients for the inversion and found stronger LD (R^2=0.50). By combining our genotype data on the inversion with genetic data from SNP chips, we successfully imputed the inversion in all 1959 ACP and 6040 control samples (info-score of the inversion =0.81, correlation of imputed inversion vs measured inversion: R^2=0.99, see online supplementary figure S6A). The major allele of the inversion conferred risk for ACP with a similar effect size as the lead SNP at this locus (OR 1.36, 95% CI 1.20 to 1.53; p=3.1×10^{-6}) (figure 3). From the ACP-associated SNPs within the *CTRB1-CTRB2* locus, *rs8048956* tagged the inversion the best (R^2=0.987) and it was therefore used as a reporter for the inversion in further analyses (see online supplementary figure S6B).

Replication of the association in NACP cohorts

To evaluate whether *rs8048956* tagging the inversion also associates with NACP, we analysed three independent cohorts of patients (n=1650) and controls (n=6695) from Germany and France (table 1 and online supplementary table S1). In all three cohorts, we found a significant association and the combined effect across all cohorts was OR 1.62 (95% CI 1.42 to 1.86); p=1.64×10^{-12} (see online supplementary figure S7A). On the other hand, association of the GWAS lead SNP *rs8055167*, which had a lower LD with the inversion, was less pronounced, supporting the pathogenic relevance of the inversion in NACP (see online supplementary figure S7B).

### Functional characterisation of the genetic rearrangement in the *CTRB1-CTRB2* locus

To investigate whether the inversion affects protein translation and secretion, CTRB1 and CTRB2 were expressed in HEK 293 T cells with the two different 5′ untranslated regions and signal peptides of the major and minor alleles. Chymotrypsin levels in the conditioned media of cells transfected with these constructs were essentially identical, indicating that the inversion has no impact on protein translation and secretion. To

Table 2 Top associated variants in the overall cohort of European alcoholic chronic pancreatitis patients

SNP ID	Locus	Gene	Counted allele	MAF	p Value	OR (95% CI)
rs545634	1p36.21	*CTRC*	A	0.12	2.8×10^{-22}	1.83 (1.63 to 2.06)
rs146437551	5q32	*SPINK1*	G	0.013	3.3×10^{-15}	3.82 (2.79 to 5.22)
rs2855983	7q34	*PRSS1-PRSS2*	G	0.392	5.5×10^{-40}	1.84 (1.69 to 2.00)
rs8055167	16q23.1	*CTRB1-CTRB2*	C	0.28	4.2×10^{-9}	1.35 (1.23 to 1.49)
rs12688091	Xq22.3	*CLDN2-MORC4*	G	0.31	9.9×10^{-33}	2.57 (2.22 to 2.98)

Results confirmed *CLDN2-MORC4*, *CTRC*, *PRSS1-PRSS2* and *SPINK1* as risk loci and identified a new locus in *CTRB1-CTRB2*. Data on linkage disequilibrium of the variants with previously reported risk variants in these loci are given in online supplementary table S3.
ID, identification; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.
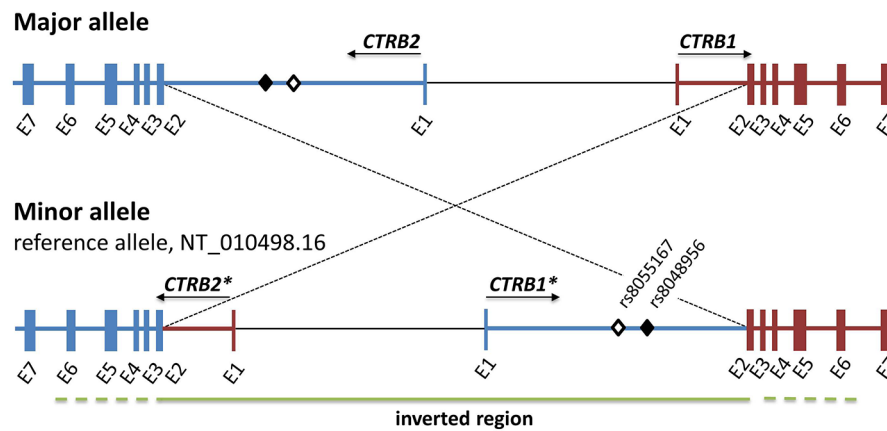
Figure 2 Schematic illustration of the *CTRB1-CTRB2* locus with the 16.6 kb inversion. The inversion breakpoints lie within the region indicated by the dashed lines. The genomic reference sequence corresponds to the minor allele. The locations of the lead single-nucleotide polymorphism (SNP) *rs8055167* and the best tagging SNP *rs8048956* are denoted by the empty and black diamond symbols, respectively. Genomic distances are not scaled. Although not shown here, in this locus *CTRB2* also harbours a 584 bp deletion variant (allele frequency ~7% in German controls) that eliminates exon 6. *CTRB1*, chymotrypsin B1 gene; *CTRB2*, chymotrypsin B2 gene; *CTRB1** and *CTRB2**, hybrid *CTRB1* and *CTRB2* genes created by the inversion; E, exon.

assess the link between the inversion and expression of *CTRB1-CTRB2*, we quantified the relative mRNA expression ratio of the two chymotrypsins using cDNA samples obtained from human pancreatic tissue or acinar cells with different inversion genotypes. We found that the major risk allele was associated with higher relative *CTRB1* expression while heterozygous carriers with one minor allele expressed higher levels of *CTRB2* (figure 4A). In accordance with our findings, data from the GTExPortal (gtexportal.org) indicated that the major allele of *rs8048956* tagging the inversion was associated with increased *CTRB1* and decreased *CTRB2* mRNA expression relative to the minor allele. The association of the lead SNP *rs8055167* with *CTRB1-CTRB2* expression was similar but smaller, supporting

a causal role of the inversion in CP (see online supplementary figure S8).

Protective degradation of anionic trypsinogen by CTRB2

When autoactivation of human anionic trypsinogen was followed at pH 8.0 in 1 mM calcium in the presence of CTRB1 or CTRB2, final trypsin levels were reduced by both chymotrypsins with a more prominent effect observed with CTRB2 (figure 4B). A similar but less pronounced effect was seen on the autoactivation of human cationic trypsinogen (PRSS1) (not shown). The stronger effect of CTRB2 was also confirmed in trypsinogen degradation experiments. SDS-PAGE analysis with Coomassie
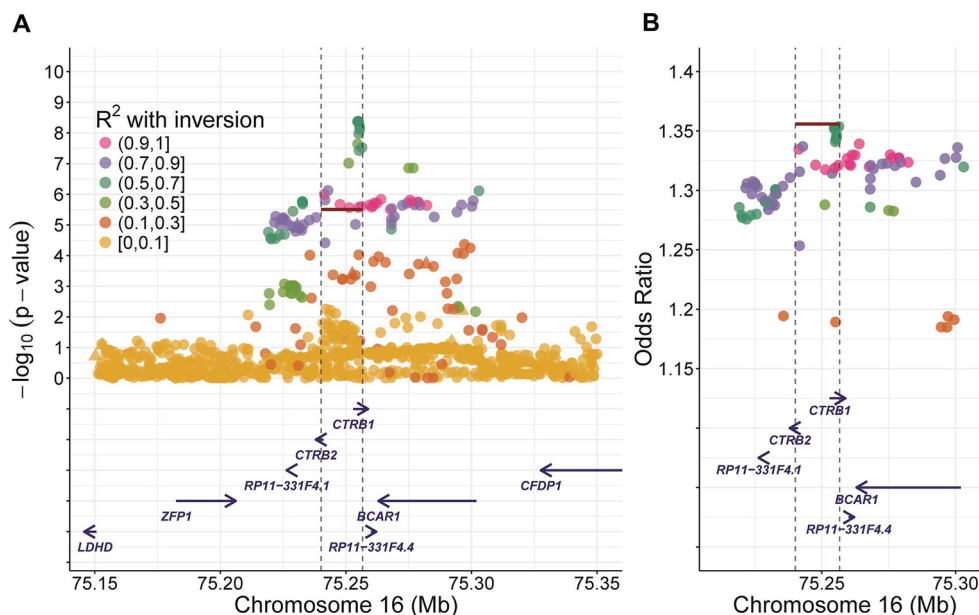
Figure 3 Regional association plots for the *CTRB1-CTRB2* locus inversion. (**A**) p Values ($-\log_{10}$) are displayed against single-nucleotide polymorphism (SNP) genomic position (genome build hg19). The inversion is represented by the red line, triangles are genotyped SNPs, circles are imputed SNPs. For calculations, all alcoholic chronic pancreatitis patients and all controls were included. (**B**) The OR-based regional association plot indicates that the association is driven by the inversion. Here ORs are represented against the genomic position.
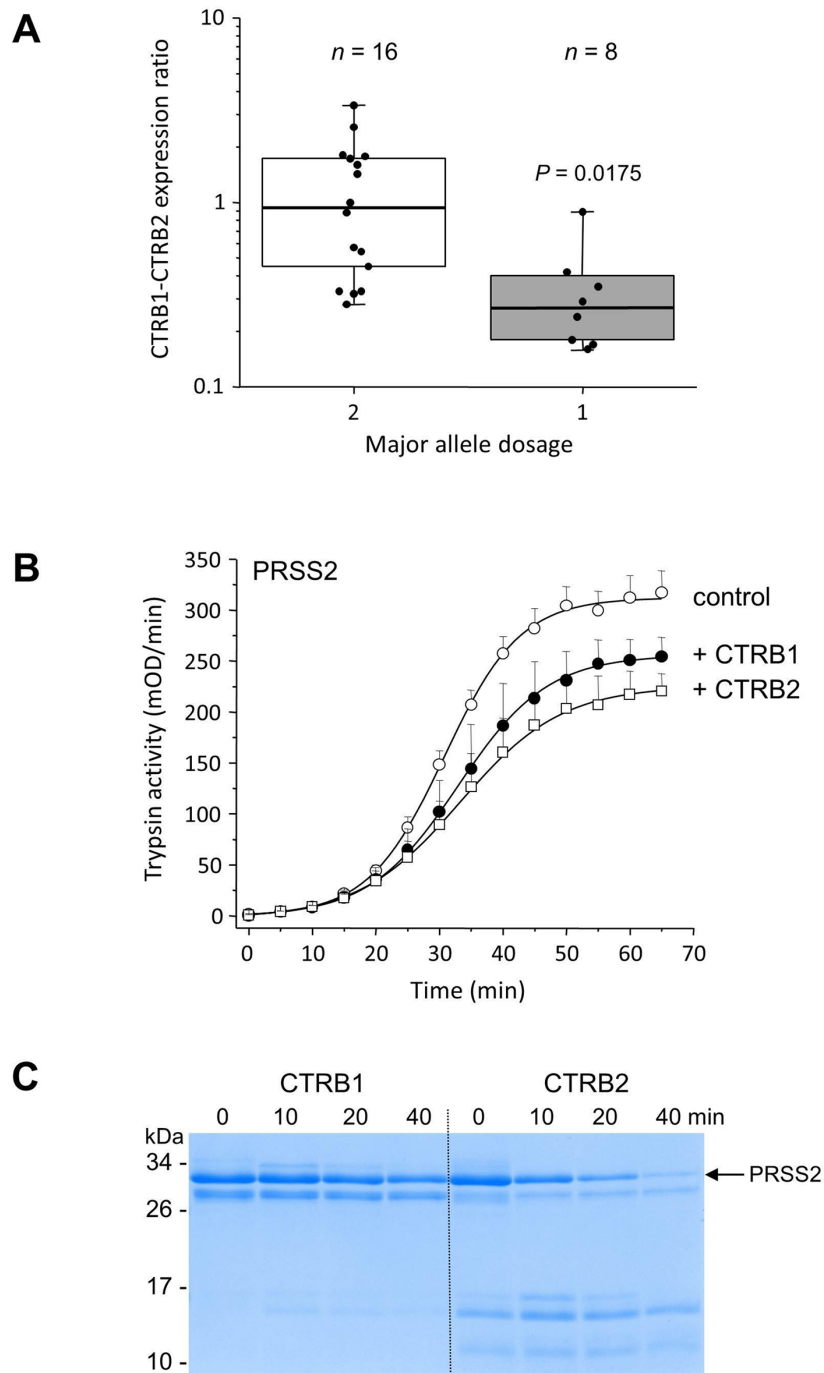
Figure 4 Expression and effect of CTRB1 and CTRB2 on trypsinogen activation and degradation. (**A**) Expression of *CTRB1* and *CTRB2* mRNA in the pancreas of subjects with different inversion genotypes. Expression ratios of *CTRB1* and *CTRB2* in pancreatic cDNA samples were determined by real-time polymerase chain reaction using the standard curve method. Results were displayed as box plots showing minimum, first quartile (25%), median, third quartile (75%), maximum and the individual values (black dots). Note that subjects carrying two copies of the major risk allele exhibit a significantly higher *CTRB1/CTRB2* expression ratio compared with heterozygous individuals with one major and one minor allele. Significance was calculated with unpaired t-test. (**B**) Effect of CTRB1 and CTRB2 on the autoactivation of human anionic trypsinogen (PRSS2). Trypsinogen (2 µM) was incubated with 10 nM initial trypsin and 200 nM of the indicated chymotrypsin in 0.1 M Tris-HCl (pH 8.0), 1 mM CaCl$_2$ and 0.05% Tween 20 (final concentrations) at 37°C in 100 µL final volume. At the indicated times, aliquots (2 µL) were withdrawn and trypsin activity was determined using 150 µM N-CBZ-Gly-Pro-Arg-p-nitroanilide substrate. Rate of substrate cleavage is given in mOD/min units measured at 405 nm. Note the lower trypsin activity that develops in the presence of chymotrypsins indicating trypsinogen degradation during activation. Under similar conditions, CTRB1 and CTRB2 had a similar but much smaller effect on the autoactivation of human cationic trypsinogen (PRSS1). (**C**) Degradation of PRSS2 by CTRB1 and CTRB2. Trypsinogen (1 µM) was incubated with 200 nM of the indicated chymotrypsin and 20 nM SPINK1 trypsin inhibitor in 0.1 M Tris-HCl (pH 8.0) and 25 mM NaCl at 37°C. Reactions were stopped at the indicated times by precipitation of 150 µL aliquots with 10% trichloroacetic acid. Samples were analysed by SDS-PAGE and Coomassie Blue staining. Note the disappearance of the intact trypsinogen band in the CTRB2 incubate. Some of the lower bands correspond to the two chains of autolysed CTRB2. Although not shown, CTRB1 or CTRB2 did not degrade cationic trypsinogen (PRSS1) to a detectable extent. CTRB1, chymotrypsin B1; CTRB2, chymotrypsin B2; PRSS2, anionic trypsinogen.

Blue staining revealed more rapid protective degradation of anionic trypsinogen (PRSS2) by CTRB2 relative to CTRB1 (figure 4C). Cationic trypsinogen (PRSS1) was not degraded by either chymotrypsin to a detectable extent (not shown).

DISCUSSION

Chronic pancreatitis is a disease of significant morbidity and suffering associated with a disproportionately large number of hospitalisations among gastrointestinal disorders. In this largest study ever conducted addressing the genetic basis of CP, we identified association with a complex genetic rearrangement in the *CTRB1-CTRB2* locus. This locus contains the two highly similar chymotrypsin genes transcribed in opposite directions. Both genes comprise seven exons and the nucleotide sequences are 97% identical with complete identity from exons 2–6. Recently, Pang *et al*[21] described a 16.6 kb inversion that exchanges the promoter region, exon 1 and intron 1 between *CTRB1* and *CTRB2*. The major, ancestral allele was arbitrarily designated as the inverted allele (figure 2). At the protein level, the inversion switches the signal peptides of CTRB1 and CTRB2, but the secreted mature proenzymes are unchanged. Furthermore, Pang *et al*[21] also reported a 584 bp deletion within the major allele leading to early termination at the protein level. Although a functional consequence of the deletion or inversion was anticipated, no disease association was reported so far.

The first novel association signal we discovered at the *CTRB1-CTRB2* locus was the lead SNP *rs8055167* located in intron 1 of *CTRB1*. This association was maintained when ACP patients were compared with alcoholics or population-based controls (see online supplementary figure S2). To understand whether the reported larger-scale genetic rearrangements might explain the association of the *CTRB1-CTRB2* locus with ACP, we analysed LD with our top SNPs. Here, an association of the 584 bp deletion was ruled out as an explanation for the association signal. However, the 16.6 kb inversion within the locus conferred risk with a similar effect size and was best tagged by *rs8048956* of the top SNPs. We also found this association in three independent European cohorts with non-alcohol-related CP using the best tagging SNP.

Functional analysis indicated that the inversion results in a reversal of the isoform expression ratio at the mRNA level which is expected to translate to a similar reversal at the protein level. Our previous studies demonstrated that primary cleavage specificity and catalytic efficiency of CTRB1 and CTRB2 are different, with CTRB2 being more active on most substrates.[31] Consequently, a change in the isoform expression ratio likely causes altered chymotrypsin activity in the pancreas.

In NACP, increased intra-pancreatic activation of trypsinogen (PRSS1) due to the failure of protective inhibition (SPINK1) and/or degradation (CTRC) is an important mechanism for disease development. The strong association of common *CTRC, PRSS1-PRSS2* and *SPINK1* locus variants in ACP indicates that this mechanism is also relevant for ACP. Therefore, we hypothesised that the altered CTRB1-CTRB2 activity profile might influence intra-pancreatic trypsinogen activation. Here, in autoactivation experiments with anionic trypsinogen, final trypsin levels were reduced by both chymotrypsins with more prominent degradation seen with CTRB2 (see figure 4). This effect for cationic trypsinogen was similar but less pronounced (not shown). Taken together, the functional studies indicate that in carriers of the major *CTRB1-CTRB2* risk allele impaired trypsinogen (PRSS2) degradation due to the altered CTRB1/CTRB2 isoform ratio explains the observed association with CP. The relatively small effect of the *CTRB1-CTRB2* inversion on CP risk is consistent with the lesser role of PRSS2 in CP (see ref.[32]).

In conclusion, our GWAS identified *CTRB1-CTRB2* as a new risk locus for ACP and NACP. The association within the *CTRB1-CTRB2* locus was linked to a 16.6 kb inversion that altered CTRB1/CTRB2 expression, thereby affecting protective trypsinogen degradation. Furthermore, we confirmed association of ACP with the *CLDN2-MORC4, CTRC, PRSS1-PRSS2* and *SPINK1* loci. Taken together, the identified risk variants explained about 18% of the variance in ACP. Our study clearly represents a significant and transformative advance in understanding the genetic basis of CP. The results underpin the prominent influence of commonly occurring variants in ACP and demonstrate that similar disease mechanisms drive both ACP and NACP. Thus, development of therapeutic approaches should be guided by disease mechanism rather than aetiology in CP.

Accession codes

The results for all imputed variants and individual-level data are available from the authors on request. The genomic reference sequence used was NT_010498.16. Reference sequences for mRNA were NM_001906.4 (CTRB1) and NM_001025200.3 (CTRB2). All reference sequences correspond to the minor inversion allele.

Author affiliations
[1]Department of Internal Medicine I, Martin Luther University, Halle, Germany
[2]Department of Internal Medicine, Neurology and Dermatology, Division of Gastroenterology, University of Leipzig, Leipzig, Germany
[3]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
[4]LIFE- Leipzig Research Center for Civilization Diseases, University of Leipzig, Germany
[5]Department of Cell Therapy, Fraunhofer Institute for Cell Therapy and Immunology (IZI), Leipzig, Germany
[6]Department of Molecular and Cell Biology, Center for Exocrine Disorders, Boston University Henry M. Goldman School of Dental Medicine, Boston, Massachusetts, USA
[7]Leipzig University Medical Center, IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany
[8]Department of Internal Medicine A, Ernst-Moritz-Arndt University, Greifswald, Germany
[9]Else Kröner-Fresenius-Zentrum für Ernährungsmedizin (EKFZ), Paediatric Nutritional Medicine, Technische Universität München (TUM), Freising, Germany
[10]Institute of Human Genetics, Helmholtz Centre Munich, German Research Centre for Environmental Health, Neuherberg, Germany
[11]Institut National de la Santé et de la Recherche Médicale (INSERM) U1078; Etablissement Français du Sang (EFS) – Bretagne; Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale; Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Régional Universitaire (CHRU) Brest, Hôpital Morvan, Brest, France
[12]Department of Gastroenterology, Technische Universität München (TUM), Munich, Germany
[13]Department of Gastroenterology & Hepatology, Erasmus Medical Centre, University Medical Centre Rotterdam, Rotterdam, The Netherlands
[14]Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service of Baden-Württemberg, Mannheim, Germany
[15]Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany
[16]Division of Gastroenterology and Gastrointestinal Endoscopy, Vita Salute San Raffaele University - San Raffaele Scientific Institute, Milan, Italy
[17]Department of Gastroenterology, Medical University of Lublin, Lublin, Poland
[18]Department of Gastroenterology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[19]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany
[20]Division of Nephrology and Dialysis, Institute of Internal Medicine, Renal Program, Columbus-Gemelli University Hospital, Catholic University, Rome, Italy
[21]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[22]Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[23]German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany
[24]Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Chirurgische Klinik, Erlangen, Germany
[25]Department of Internal Medicine I, University Hospital Regensburg, Regensburg, Germany
[26]Institute for Translational Medicine and First Department of Internal Medicine, University of Pécs, Pécs, Hungary
[27]HAS-SZTE, Momentum Gastroenterology Multidisciplinary Research Group, Szeged, Hungary
[28]Department of Internal Medicine and Gastroenterology, University of Medicine and Pharmacy, Craiova, Romania
[29]Department of Gastroenterology and Internal Medicine, Medical University Bialystok, Bialystok, Poland
[30]Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany
[31]TIB MOLBIOL, Berlin, Germany
[32]Pôle des Maladies de l'Appareil Digestif, Service de Gastroentérologie-Pancréatologie, Hôpital Beaujon, AP-HP, Clichy, France
[33]Gastrocentrum, Karolinska Institutet CLINTEC, Stockholm, Sweden
[34]Department of Biology and Medical Genetics, University Hospital Motol and 2nd Faculty of Medicine, Charles University, Prague, Czech Republic
[35]Grupo de Epidemiología Genética y Molecular Programa de Genética del Cáncer Humano Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain
[36]CIBERONC, Spain
[37]Department of Digestive Tract Diseases, Medical University of Łódź, Łódź, Poland
[38]Biology and Genetics, Department of Life and Reproduction Sciences, University of Verona, Verona, Italy
[39]Department of Medicine II, University Hospital,Ludwig-Maximilians-University Munich, Munich, Germany
[40]Department of Gastroenterology and Hepatology, Radboud umc, Nijmegen, The Netherlands
[41]Department of Internal Medicine, Salem Medical Centre and Centre for Alcohol Research, University of Heidelberg, Heidelberg, Germany
[42]Department of Pediatrics I, Medical University, Innsbruck, Austria
[43]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
[44]Institute of Human Genetics, University of Bonn, Bonn, Germany
[45]Department of Medical and Surgical Sciences, IV Surgical Clinic, University of Padua, Padua, Italy
[46]Division of Medicine, UCL Institute for Liver and Digestive Health, University College London, London, UK
[47]DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
[48]Clinic for Internal Medicine, Hospital Döbeln, Döbeln, Germany
[49]Epithelial Carcinogenesis Group, Molecular Pathology Programme, Centro Nacional de Investigaciones Oncológicas, Madrid, Spain
[50]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain
[51]Department of Psychiatry, University of Regensburg, Regensburg, Germany
[52]Department of Gastroenterology, Hepatology, Infectious Diseases, Medical Faculty of Mannheim University of Heidelberg, Mannheim, Germany
[53]Department of Surgery, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
[54]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
[55]Department of Haematology, University of Cambridge, Cambridge, UK
[56]Psychiatric Hospital, University of Munich, Munich, Germany
[57]Department of Surgery and Interventional Science, University College London, London, UK
[58]Department of Gastroenterology and Hepatology, University Hospital of Zürich, Zürich, Switzerland
[59]Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[60]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[61]Department of Internal Medicine, Neurology and Dermatology, Division of Endocrinology, University of Leipzig, Leipzig, Germany
[62]Department of General, Visceral, and Transplant Surgery, Ludwig Maximilian University, Munich, Germany
[63]Division of Gastroenterology, Tohoku University Graduate School of Medicine, SendaiMiyagi, Japan

Competing interests None declared.

Patient consent Obtained.

Ethics approval Ethic Committee of the University of Leipzig.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All chip data are available upon request.

REFERENCES

1 Witt H, Apte MV, Keim V, et al. Chronic pancreatitis: challenges and advances in pathogenesis, genetics, diagnosis, and therapy. *Gastroenterology* 2007;132:1557–73.
2 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.
3 Witt H, Luck W, Hennies HC, et al. Mutations in the gene encoding the serine protease inhibitor, kazal type 1 are associated with chronic pancreatitis. *Nat Genet* 2000;25:213–6.
4 Le Maréchal C, Masson E, Chen JM, et al. Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* 2006;38:1372–4.
5 Rosendahl J, Witt H, Szmola R, et al. Chymotrypsin C (*CTRC*) variants that diminish activity or secretion are associated with chronic pancreatitis. *Nat Genet* 2008;40:78–82.
6 Witt H, Luck W, Becker M, et al. Mutation in the *SPINK1* trypsin inhibitor gene, alcohol use, and chronic pancreatitis. *JAMA* 2001;285:2716–7.
7 Schnúr A, Beer S, Witt H, et al. Functional effects of 13 rare PRSS1 variants presumed to cause chronic pancreatitis. *Gut* 2014;63:337–43.

8 Witt H, Beer S, Rosendahl J, et al. Variants in CPA1 are strongly associated with early onset chronic pancreatitis. Nat Genet 2013;45:1216–20.

9 Lankisch PG, Lowenfels AB, Maisonneuve P. What is the risk of alcoholic pancreatitis in heavy drinkers? Pancreas 2002;25:411–2.

10 Kristiansen L, Grønbaek M, Becker U, et al. Risk of pancreatitis according to alcohol drinking habits: a population-based cohort study. Am J Epidemiol 2008;168:932–7.

11 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.

12 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.

13 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.

14 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.

15 Giri AK, Midha S, Banerjee P, et al. Common variants in CLDN2 and MORC4 genes Confer Disease susceptibility in patients with chronic pancreatitis. PLoS One 2016;11:e0147345.

16 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.

17 Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.

18 Price AL, Weale ME, Patterson N, et al. Long-range LD can confound genome scans in admixed populations. Am J Hum Genet 2008;83:132–5.

19 Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.

20 de Bakker PI, Ferreira MA, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008;17:R122–R128.

21 Pang AW, Migita O, Macdonald JR, et al. Mechanisms of formation of structural variation in a fully sequenced human genome. Hum Mutat 2013;34:345–54.

22 Loeffler M, Engel C, Ahnert P, et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. BMC Public Health 2015;15:691.

23 Király O, Guan L, Sahin-Tóth M. Expression of recombinant proteins with uniform N-termini. Methods Mol Biol 2011;705:175–94.

24 Szabó A, Pilsak C, Bence M, et al. Complex formation of human proelastases with Procarboxypeptidases A1 and A2. J Biol Chem 2016;291:17706–16.

25 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.

26 Paliwal S, Bhaskar S, Mani KR, et al. Comprehensive screening of chymotrypsin C (CTRC) gene in tropical calcific pancreatitis identifies novel variants. Gut 2013;62:1602–6.

27 Masamune A, Nakano E, Kume K, et al. Identification of novel missense CTRC variants in Japanese patients with chronic pancreatitis. Gut 2013;62:653.2–4.

28 LaRusch J, Lozano-Leon A, Stello K, et al. The common chymotrypsinogen C (CTRC) Variant G60G (C.180T) Increases risk of chronic pancreatitis but not recurrent acute Pancreatitis in a North American Population. Clin Transl Gastroenterol 2015;6:e68.

29 Aoun E, Chang CC, Greer JB, et al. Pathways to injury in chronic pancreatitis: decoding the role of the high-risk SPINK1 N34S haplotype using meta-analysis. PLoS One 2008;3:e2003.

30 Hurley TD, Edenberg HJ. Genes encoding enzymes involved in ethanol metabolism. Alcohol Res 2012;34:339–44.

31 Szabó A, Sahin-Tóth M. Determinants of chymotrypsin C cleavage specificity in the calcium-binding loop of human cationic trypsinogen. Febs J 2012;279:4283–92.

32 Jancsó Z, Sahin-Tóth M. Tighter control by chymotrypsin C (CTRC) Explains lack of Association between human anionic trypsinogen and hereditary pancreatitis. J Biol Chem 2016;291:12897–905.