Pancreatology

Colocalization analysis of pancreas eQTLs with risk loci from alcoholic and novel non-alcoholic chronic pancreatitis GWAS suggests potential disease causing mechanisms

--Manuscript Draft--

Manuscript Number:	PANCREATOLOGY-D-22-00076
Article Type:	Original Article
Section/Category:	Chronic pancreatitis
Keywords:	Bayesian colocalization analysis; eQTL; GWAS; Chronic pancreatitis; GTEx
Corresponding Author:	Andreas W. Schmidt Martin-Luther-University Halle-Wittenberg Halle (Saale), GERMANY
First Author:	Andreas W. Schmidt
Order of Authors:	Andreas W. Schmidt
	Andreas Kühnapfel
	Holger Kirsten
	Harald Grallert
	Claus Hellerbrand
	Falk Kiefer
	Karl Mann
	Sebastian Mueller
	Markus M. Nöthen
	Annette Peters
	Monika Ridinger
	Josef Frank
	Marcella Rietschel
	Nicole Soranzo
	Michael Soyka
	Norbert Wodarz
	Giovanni Malerba
	Giovanni Gambaro
	Christian Gieger
	Markus Scholz
	Sebastian Krug
	Patrick Michl
	Maren Ewers
	Heiko Witt
	Helmut Laumen
	Jonas Rosendahl
Abstract:	Background/Objectives Previous genome-wide association studies (GWAS) identified genome-wide significant risk loci in chronic pancreatitis and investigated underlying disease causing mechanisms by simple overlaps with expression quantitative trait loci

	(eQTLs), a procedure which may often result in false positive conclusions. Methods We conducted a GWAS in 584 non-alcoholic chronic pancreatitis (NACP) patients and 6040 healthy controls. Next, we applied Bayesian colocalization analysis of identified genome-wide significant risk loci from both, our recently published alcoholic chronic pancreatitis (ACP) and the novel NACP dataset, with pancreas eQTLs from the GTEx V8 European cohort to prioritize candidate causal genes and extracted credible sets of shared causal variants. Results Variants at the CTRC (p = 1.22 x 10-21) and SPINK1 (p = 6.59 x 10-47) risk loci reached genome-wide significance in NACP. CTRC risk variants colocalized with CTRC eQTLs in ACP (PP4 = 0.99, PP4/PP3 = 95.51) and NACP (PP4 = 0.99, PP4/PP3 = 95.46). For both diseases, the 95% credible set of shared causal variants consisted of rs497078 and rs545634. CLDN2-MORC4 risk variants colocalized with CLDN2 eQTLs in ACP (PP4 = 0.98, PP4/PP3 = 42.20) and NACP (PP4 = 0.67, PP4/PP3 = 7.18), probably driven by the shared causal variant rs12688220. Conclusions A shared causal CTRC risk variant might unfold its pathogenic effect in ACP and NACP by reducing CTRC expression, while the CLDN2-MORC4 shared causal variant rs12688220 may modify ACP and NACP risk by increasing CLDN2 expression.
Suggested Reviewers:	Matthias Löhr matthias.lohr@ki.se
	Atsushi Masamune amasamune@med.tohoku.ac.jp
	Eszter Hegyi eszter.hegyi@aok.pte.hu
	Giovanni Marchegiani giovanni.marchegiani@univr.it
Opposed Reviewers:	David C. Whitcomb whitcomb@pitt.edu Conflict of interest

Universitätsklinikum Halle (Saale)

Universitätsklinikum Halle (Saale) | Postfach | 06097 Halle (Saale)

Mitglied im: KKH Krukenberg Krebszentrum Halle

Department Innere Medizin

Universitätsklinik und Poliklinik für Innere Medizin I Direktor: Prof. Dr. med. Patrick Michl

To the

Editor-in-Chief of
Pancreatology
Dr. Miklós Sahin-Tóth

Halle, 11th February 2022

Hausanschrift:
Ernst-Grube-Straße 40
06120 Halle (Saale)

innere1@uk-halle.de
www.medizin.uni-halle.de/kim1

Direktionssekretariat
Telefon: 0345 557-2661
Telefax: 0345 557-2253

Oberarztsekretariat
Telefon: 0345 557-2665

Leiter Pneumologie
Dr. Stephan Eisenmann Tel. -3239

Sekretariat Pneumologie
Telefon: 0345 557-3238

Endoskopie Tel. -2771
Ultraschall Tel. -2187
Gastroenterologisches Funktionslabor Tel. -2672

Stationen
Gastroenterologie St. 6 Tel. -2645
Pneumologie St. 9 Tel. -2640
IKL Tel. -3047

Ambulanzen **Tel. -3392**
Hepatologie/ Lebertransplantation
Morbus Crohn / Colitis ulcerosa
Pneumologie
Pulmonale Hypertonie
Mukoviszidose

Gastroent. / pneum. Onkologie
Tel. -7233

Studienzentrum
Tel. -2668

Forschung
Molekulare Hepatologie
Tel. -2138
Gastrointestinale Onkologie
Tel. -2138
Pankreatologie
Tel. -2138
Molekulare Pneumologie
Tel. -7013

Dear Dr. Miklós Sahin-Tóth,

Please find attached our manuscript entitled "**Colocalization analysis of pancreas eQTLs with risk loci from alcoholic and novel non-alcoholic chronic pancreatitis GWAS suggests potential disease causing mechanisms**" which we would like to submit for consideration for publication in *Pancreatology* as an "**Original article**".

Chronic pancreatitis (CP) is one of the most frequent gastrointestinal diagnoses requiring hospitalization of patients. Thus far, genome-wide association studies (GWAS) revealed several common risk loci that contribute to CP development. Here, we conducted a GWAS analysis of non-alcoholic chronic pancreatitis (NACP) cases and found genome-wide significance for the *SPINK1* and the *CTRC* locus.

Next, we aimed to discover molecular mechanisms at the discovered NACP risk loci and at risk loci from our previously published alcoholic chronic pancreatitis (ACP) GWAS, by investigation of expression quantitative trait loci (eQTLs) from the European GTEx pancreas dataset. We identified eight potential regulated genes, which were further prioritized with Bayesian colocalization analysis. In ACP and NACP, we found evidence of colocalization between *CTRC* and *CLDN2-MORC4* GWAS signals and *CTRC* and *CLDN2* eQTLs, respectively. Finally, we present the potentially shared causal variants behind the CP associations and eQTL signals.

We believe that our study is of interest to the readership of *Pancreatology* as it supports well established CP risk loci, gives mechanistic explanations at known CP

risk loci and furthermore introduces the statistical method of Bayesian colocalization to the pancreatitis community.

We thank you for your consideration and look forward to hearing about the disposition of our manuscript in the near future.

Yours sincerely,

Jonas Rosendahl, MD

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Colocalization analysis of pancreas eQTLs with risk loci from alcoholic and novel non-alcoholic chronic pancreatitis GWAS suggests potential disease causing mechanisms

Andreas W. Schmidt[1,2], Andreas Kühnapfel[3,4], Holger Kirsten[3,4], Harald Grallert[5-7], Claus Hellerbrand[8], Falk Kiefer[9], Karl Mann[9], Sebastian Mueller[10], Markus M. Nöthen[11-12], Annette Peters[6,7,13], Monika Ridinger[14], Josef Frank[15], Marcella Rietschel[15], Nicole Soranzo[16-17], Michael Soyka[18], Norbert Wodarz[14], Giovanni Malerba[19], Giovanni Gambaro[20], Christian Gieger[5-7], Markus Scholz[3,4] Sebastian Krug[1], Patrick Michl[1], Maren Ewers[2], Heiko Witt[2], Helmut Laumen[1#*], Jonas Rosendahl[1#*]

[#] These authors contributed equally and share last authorship, * corresponding authors

[1] Department of Internal Medicine I, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

[2] Else Kröner-Fresenius-Zentrum für Ernährungsmedizin (EKFZ), Paediatric Nutritional Medicine, Technische Universität München (TUM), Freising, 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, Leipzig, Germany

[5] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[6] Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[7] German Center for Diabetes Research (DZD e.V.), Neuherberg, Germany

[8] Institute of Biochemistry, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

[9] Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany

[10] Department of Internal Medicine, Salem Medical Centre and Centre for Alcohol Research, University of Heidelberg, Heidelberg, Germany

[11] Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany

[12] Institute of Human Genetics, University of Bonn, Bonn, Germany

[13] Institute for Medical Information Processing Biometry and Epidemiology (IBE), Ludwig-Maximilians-Universität München, Munich, Germany

[14] Department of Psychiatry, University of Regensburg, Regensburg, Germany

[15] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany

[16] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[17] Department of Haematology, University of Cambridge, Cambridge, UK

[18] Psychiatric Hospital, University of Munich, Munich, Germany

[19] Biology and Genetics, Department of Life and Reproduction Sciences, University of Verona, Verona, Italy

[20] Division of Nephrology, Department of Medicine, University of Verona, Italy

***Corresponding authors:**

Jonas Rosendahl, MD

Department of Medicine; Clinic for Internal Medicine

Martin-Luther-University Halle-Wittenberg

Ernst-Grube-Str. 40

D-06120 Halle (Saale)

Germany

E-Mail: jonas.rosendahl@uk-halle.de

Phone: +49 345 557 2661

Fax: +49 345 557 2653

ORCID ID: 0000-0002-5885-5242

Helmut Laumen, PhD

Department of Medicine; Clinic for Internal Medicine

Martin-Luther-University Halle-Wittenberg

Ernst-Grube-Str. 40

D-06120 Halle (Saale)

Germany

E-Mail: helmut.laumen@medizin.uni-halle.de

Phone:+49 345 557 1241

Fax: +49 345 557 2653

ORCID ID: 0000-0003-4513-0506

Running title:

Colocalization of pancreas eQTLs with pancreatitis loci

Key words:

Bayesian colocalization analysis, eQTL, GWAS, chronic pancreatitis, GTEx

Disclosure statement:

The authors report no conflicts of interest.

Financial support:

The authors thank all study participants for providing clinical data and blood samples. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) grants RO3929/4-1, RO3929/5-1, RO3929/6-1 & RO3929/7-1 (to J.R.), by the Else Kröner-Fresenius-Foundation (EKFS) 2017_A108 – EKFZ Witt (to H.W.).

Author contributions:

A.W.S., H.W., H.L. and J.R. conceived and directed the study.

A.K., and H.K. conducted GWAS analysis.

A.W.S. performed genetic data collection and statistic/bioinformatic analysis.

A.W.S., H.L., and J.R. drafted and revised the manuscript with substantial help from H.W.
All other co-authors recruited study subjects, collected clinical data and/or provided genomic DNA samples. All authors approved the final manuscript and contributed critical revisions to its intellectual content.

List of Abbreviations:

CP, chronic pancreatitis; NACP, non-alcoholic chronic pancreatitis; ACP, alcoholic chronic pancreatitis; HGNC, HUGO Gene Nomenclature Committee; GWAS, Genome-wide association study; eQTL, expression quantitative trait locus; TSS, transcriptional start site; LD, linkage disequilibrium; OR, Odds ratio; CI, confidence interval

Abstract

Background/Objectives Previous genome-wide association studies (GWAS) identified genome-wide significant risk loci in chronic pancreatitis and investigated underlying disease causing mechanisms by simple overlaps with expression quantitative trait loci (eQTLs), a procedure which may often result in false positive conclusions.

Methods We conducted a GWAS in 584 non-alcoholic chronic pancreatitis (NACP) patients and 6040 healthy controls. Next, we applied Bayesian colocalization analysis of identified genome-wide significant risk loci from both, our recently published alcoholic chronic pancreatitis (ACP) and the novel NACP dataset, with pancreas eQTLs from the GTEx V8 European cohort to prioritize candidate causal genes and extracted credible sets of shared causal variants.

Results Variants at the *CTRC* ($p = 1.22 \times 10^{-21}$) and *SPINK1* ($p = 6.59 \times 10^{-47}$) risk loci reached genome-wide significance in NACP. *CTRC* risk variants colocalized with *CTRC* eQTLs in ACP (PP4 = 0.99, PP4/PP3 = 95.51) and NACP (PP4 = 0.99, PP4/PP3 = 95.46). For both diseases, the 95% credible set of shared causal variants consisted of *rs497078* and *rs545634*. *CLDN2-MORC4* risk variants colocalized with *CLDN2* eQTLs in ACP (PP4 = 0.98, PP4/PP3 = 42.20) and NACP (PP4 = 0.67, PP4/PP3 = 7.18), probably driven by the shared causal variant *rs12688220*.

Conclusions A shared causal *CTRC* risk variant might unfold its pathogenic effect in ACP and NACP by reducing *CTRC* expression, while the *CLDN2-MORC4* shared causal variant *rs12688220* may modify ACP and NACP risk by increasing *CLDN2* expression.

Introduction

Chronic pancreatitis (CP) is a recurring or continuous inflammation of the pancreas with exocrine and endocrine impairment of function. The predominant aetiology is alcohol abuse, whereas other factors are less common. In recent years, however, targeted and untargeted genetic studies have identified several genetic risk loci in alcoholic chronic pancreatitis (ACP) and non-alcoholic chronic pancreatitis (NACP) patients. So far, genetic defects can be assigned to a trypsin-centred and a ductal pathway in most cases [1].

The first risk gene discovered in CP was *PRSS1*, encoding the digestive enzyme cationic trypsinogen [2]. Coding mutations in this gene increase trypsin activity [3], which results in autodigestion of the pancreas. Further studies identified mutations in the genes *SPINK1* and *CTRC* [4,5], encoding a trypsin inhibitor and a trypsin degrading protease, respectively. In addition, candidate studies discovered variants in further genes, encoding the digestive enzymes *PNLIP*, *CEL* and *CPA1* [6–8] and the pancreatic ductal membrane proteins *CFTR* and *TRPV6* [9,10].

The first genome-wide association study (GWAS) in CP revealed genome-wide significant risk variants at the *PRSS1-PRSS2* and *CLDN2-MORC4* loci and investigated potential disease causing mechanisms. Only for the *PRSS1-PRSS2* risk locus, a significant correlation between risk genotype and expression levels of a neighbouring gene (*PRSS1*) was reported, whereas the risk allele at the *CLDN2-MORC4* locus has been shown to be associated with an atypical localization of claudin-2 [11]. More recent association studies replicated both, the *PRSS1-PRSS2* and the *CLDN2-MORC4* locus [12–14]. Furthermore, genome-wide significant associations for the risk loci *CTRB1-CTRB2*, *CTRC* and *SPINK1* were identified in a recent ACP GWAS [13]. At the *CTRB1-CTRB2* locus, the major allele of the SNP *rs8048956* was linked to a 16.6 kb inversion, which might contribute to CP risk by modulating the gene expression levels of *CTRB1* and *CTRB2*, leading to reduced anionic trypsinogen degradation [13].

Both previous CP GWAS studies [11,13] investigated disease causing mechanisms by simple overlaps with eQTL data, a procedure which can be misleading as causal variants might be different. To uncover the biological mechanisms behind GWAS loci, refined analysis techniques are required. Bayesian colocalization analysis is a prominent approach to infer shared causal variants from association datasets, indicating disease causing mechanisms [15]. In the present study, we applied a GWAS to identify genome-wide significant risk loci in NACP. Next, we investigated colocalization of the here identified risk loci and recently published ACP risk loci [13] with pancreas specific European public domain eQTL data [16].

We found supportive evidence for causal genes and prioritized credible sets of shared causal variants at the respective ACP and NACP risk loci.

Methods

Study population. The study cohorts consisted of 584 NACP cases, 1959 ACP cases and 6040 controls, all of European ancestry (**Supplementary Table 1**) [13]. Chronic pancreatitis was defined according to recent guidelines, if patients showed recurrent attacks or chronic pain, characteristic morphological changes in imaging studies and functional impairment of exocrine/endocrine pancreas function [17]. ACP was diagnosed when alcohol consumption was > 60g in females and > 80g in males per day for at least 2 years [13]. All included subjects gave their informed consent and in all participating study centres the corresponding medical ethical review committees approved the study.

ACP and NACP GWAS-analysis. Genotyping and imputation was described in Rosendahl et al. [13]. Variants were filtered for MAF ≥ 0.01 and imputation info score ≥ 0.5 in ACP and NACP, respectively. Logistic regression of ACP and NACP datasets was conducted in PLINK 2.0 [18] with the first three principal components of the variant data included as covariates. Genomic control was applied to standard errors and p-values [19]. A Manhattan plot of the novel NACP GWAS dataset was generated with the function "manhattan" from the R package "qqman" [20].

Pre-filtering of eQTL genes. In brief, a set of input eQTL genes for the subsequent colocalization analysis was inferred by overlapping top GWAS variants at CP risk loci with significant variant-gene pairs from pancreas tissue (downloaded from https://www.gtexportal.org/home/datasets). Top GWAS variants were defined as the respective lead variants or variants in high LD ($r^2 \geq 0.7$, max. 250 kb distance). For calculations of LD, PLINK 1.9 [18] was used, applying the 1000Genomes European reference phase 1, V. 3 panel [21] as reference population. For lead variants that were not present in the 1000Genomes European cohort, the variant with the second or next strongest association at the respective locus was considered to define the set of high-LD variants, including the initial lead variant for further analysis. To merge both datasets, the GWAS variants were first lifted from hg19 to hg38, using the UCSC chain file (https://hgdownload.soe.ucsc.edu/gbdb/hg19/liftOver/) applying the R function "liftover" from the R package "rtracklayer" [22]. An eQTL gene was considered for Bayesian colocalization analysis, when sharing a significant eQTL signal in the pancreas with a top GWAS variant. Ensemble gene IDs were translated to HGNC symbols with the R package "biomaRt" [23].

Bayesian colocalization analysis. The applied Bayesian colocalization analysis investigates whether two association signals are consistent with a single shared causal variant [15]. In the case of a GWAS and an eQTL dataset, for each locus the posterior probabilities (**PP0** – **PP4**) were calculated for five hypotheses; **H0**: *No association within either GWAS or eQTL dataset*, **H1**: *Association only within GWAS dataset*, **H2**: *Association*

only within eQTL dataset, **H3**: *Association within GWAS and eQTL dataset, but independent variants,* and **H4**: *Shared causal variant associated within GWAS and eQTL dataset*; based on the input data and defined prior probabilities $\mathbf{p_1}$: *Prior probability for associated variant within GWAS dataset*, $\mathbf{p_2}$: *Prior probability for associated variant within eQTL dataset*, and $\mathbf{p_{12}}$: *Prior probability for shared associated variant*.

Full European pancreas cis-eQTL summary statistics were downloaded from the GTEx V8 google cloud storage repository (https://console.cloud.google.com/storage/browser/gtex-resources). Bayesian colocalization analysis was applied on our NACP/ACP GWAS (lifted to hg38) and public domain European GTEx pancreas eQTLs with the "coloc.abf" function from the R package "coloc" [15], using beta coefficients (effect size) and standard errors (squared to obtain variances) as inputs. For the eQTL dataset, we furthermore defined the "sdY" parameter to be 1 since GTEx expression values for each gene were normalized across samples using an inverse normal transform [16]. The default value 1×10^{-4} was chosen for the coloc parameters $\mathbf{p_1}$ and $\mathbf{p_2}$, for $\mathbf{p_{12}}$ a stringent parameter setting of 1×10^{-6} was chosen [15]. Coloc results of PP4 > 0.5 and PP4/PP3 ≥ 5 were considered as sufficient evidence of colocalization [24]. Bayesian colocalization analysis was conducted including all variants at CP risk loci annotated in both, the GWAS and eQTL datasets, ±1 MB from the transcriptional start site (TSS) of a given eQTL gene (number of input variants in **Supplementary Table 2**).

To further evaluate the colocalization results, we visually inspected scatter plots of $-\log_{10}$eQTL/$-\log_{10}$GWAS p-values for each investigated eQTL gene with the R package "ggplot2" [25]. A single spike of data points towards the top right corner and also top right data points may support colocalization [26]. Distinct (clusters of) top variants biased towards both, eQTL and GWAS signal strength (highlighted in red and green, respectively) speak against colocalization. At each colocalizing locus (PP4 > 0.5 and PP4/PP3 ≥ 5), we extracted the 95% credible set of shared causal variants from colocalization result objects in R, calculated from their respective posterior probabilities to be shared causal ("SNP.PP.H4").

Results

Identification of genome-wide significant NACP risk loci

After conducting a GWAS analysis in 584 NACP cases and 6.040 controls, we discovered the two already reported CP risk loci *SPINK1* and *CTRC* reaching genome-wide significance (**Figure 1, Supplementary Table 3**). The here identified *SPINK1* lead variant (*rs17107296*, $p = 6.59 \times 10^{-47}$, OR = 11.25, CI = 8.09 to 15.64, **Supplementary Table 3**) was in perfect LD ($r^2 = 1.0$) with the known pancreatitis risk mutation p.N34S [4] (*rs17107315*, $p = 3.88 \times 10^{-46}$, OR = 10.97, CI = 7.89 to 15.25, **Supplementary Table 3**). Notably, the association was stronger when compared to ACP (*rs17107296*: $p = 4.15 \times 10^{-15}$, OR = 3.81, CI = 2.73 to 5.31, **Supplementary Table 5**). The *CTRC* lead variant *rs497078* ($p = 1.22 \times 10^{-21}$, OR = 2.28, CI =1.93 to 2.70, **Supplementary Table 3**) was in high LD ($r^2 = 0.85$) with our previously discovered ACP lead variant *rs545634* ($p = 1.97 \times 10^{-20}$, OR = 2.24, CI = 1.89 to 2.66, **Supplementary Table 3**).

Selection of input eQTL genes for Bayesian colocalization analysis

After identification of genome-wide significant risk loci in NACP, we hypothesized that risk variants in CP may unfold their pathogenic potential by causing aberrant gene expression of nearby genes. Since known CP risk genes are predominantly expressed in the pancreas, we mapped pancreas eQTL data from the GTEx European cohort to our ACP and NACP GWAS datasets.

We overlapped top GWAS variants at risk loci (lead variants and variants $r^2 \geq 0.7$ within a 250 kb window) with significant pancreas eQTLs ("significant variant-gene pairs" from GTEx V8). For all significant pancreas eQTLs with a top GWAS variant overlap, the respective gene was considered for the following Bayesian colocalization analysis. Note that the ACP lead variant *rs2855983* was not annotated in the 1000Genomes European reference, so we obtained variants in high LD with the second strongest signal at this locus *rs4726575* to form the set of top variants.

In both, NACP and ACP, we obtained the eQTL gene *CTRC* at the *CTRC* risk locus (**Supplementary Tables 4 and 5**). In ACP we obtained the eQTL genes *CLDN2* at the *CLDN2-MORC4* risk locus, *CTRB1*, *CTRB2*, ENSG00000240338 and ENSG00000280152 at the *CTRB1-CTRB2* risk locus and *PRSS2* and *TRBV29-1* at the *PRSS1-PRSS2* risk locus (**Supplementary Table 5**). The liftover from hg19 to hg38 of the ACP lead variant *rs2855983* failed. Therefore, we manually integrated the *rs2855983* GTEx annotations for significant *PRSS2* and

TRBV29-1 eQTLs in the following colocalization analysis (**Supplementary Table 5**). Note, that all top *SPINK1* variants could not be mapped to any significant pancreas eQTLs (**Supplementary Tables 4 and 5**).

Identification of candidate causal genes using Bayesian colocalization analysis

To further prioritize the obtained eQTL genes, we conducted Bayesian colocalization analysis at the genome-wide significant risk locus *CTRC* in NACP and the *CTRC*, *CLDN2-MORC4*, *CTRB1-CTRB2* and *PRSS1-PRSS2* genome-wide significant risk loci in ACP. Assuming similarities between ACP and NACP, we also analysed colocalization for the ACP inferred risk loci *CLDN2-MORC4*, *CTRB1-CTRB2 and PRSS1-PRSS2* in the NACP cohort with the discovered eQTL genes from ACP. Furthermore, we plotted the $-\log_{10}$eQTL/$-\log_{10}$GWAS p-values for visual investigation (**Figure 2** and **Figure 3**).

We discovered evidence of colocalization between *CLDN2-MORC4* risk variants and *CLDN2* eQTLs in ACP and NACP (ACP: PP4 = 0.98, PP4/PP3 = 42.20; NACP: PP4 = 0.67, PP4/PP3 = 7.18; **Table 1**). Supporting this observation, the respective scatter plots of eQTL and GWAS signals formed single spikes towards the top right corner (**Figure 2A and Figure 3A**).

For GWAS risk variants at the *CTRB1-CTRB2* locus, we found no evidence of colocalization with *CTRB1* or *CTRB2* eQTLs in ACP (*CTRB1* PP4 = 0.13, PP4/PP3 = 0.16; *CTRB2* PP4 = 0.04, PP4/PP3 = 0.04; **Table 1**). The respective scatter plots indicated the presence of distinct top clusters of variants, biased either towards eQTL or GWAS association (**Figure 2B-C**, highlighted in red and green). In the corresponding NACP scatter plots, we observed single spikes towards the top right corner (**Figure 3B-C**), but Bayesian colocalization analysis revealed no evidence of colocalization with *CTRB1* or *CTRB2* eQTLs (*CTRB1* PP4 = 0.49, PP4/PP3 = 1.93; *CTRB2* PP4 = 0.49, PP4/PP3 = 1.94; **Table 1**). The eQTL gene ENSG00000240338 at this locus showed no evidence of colocalization in ACP or NACP (ACP: PP4 = 0.70, PP4/PP3 = 2.31; NACP: PP4 = 0.04, PP4/PP3 = 0.09; **Table 1**) and the respective scatter plots indicated distinct top clusters of variants, biased either towards eQTL or GWAS association (**Figure 2E and Figure 3E,** highlighted in red and green). Similarly, eQTLs for the gene ENSG00000280152 at the *CTRB1-CTRB2* locus did not colocalize with GWAS association signals in ACP or NACP (ACP: PP4 = 0.07, PP4/PP3 = 0.07; NACP: PP4 = 0.33, PP4/PP3 = 0.99; **Table 1**) and the scatter plots showed distinct clusters of top variants for eQTL and GWAS association (**Figure 2F and Figure 3F**, highlighted in red and green).

At the *CTRC* risk locus, we discovered colocalization of ACP and NACP risk variants with *CTRC* eQTLs (ACP:

PP4 = 0.99, PP4/PP3 = 95.51; NACP: PP4 = 0.99, PP4/PP3 = 95.46; **Table 1**). Both respective scatter plots contained top right data points (**Figure 2D and 3D**).

Risk variants at the *PRSS1-PRSS2* risk locus showed no evidence of colocalization in ACP or NACP with *PRSS2* eQTLs (ACP: PP4 = 0.64, PP4/PP3 = 2.57; NACP: PP4 = 0.36, PP4/PP3 = 0.91; **Table 1**). Both respective scatter plots were also not indicative of colocalization (**Figure 2G and 3G**). Furthermore, *PRSS1-PRSS2* risk variants were also not colocalizing with *TRBV29-1* eQTLs in ACP or NACP (ACP: PP4 = 0.00, PP4/PP3 = 0.00; NACP: PP4 = 0.58, PP4/PP3 = 1.59; **Table 1**). The *TRBV29-1* scatter plots in ACP and NACP showed distinct top variants biased either towards eQTL or GWAS association (**Figure 2H and Figure 3H**, highlighted in red and green).

Statistical fine-mapping of shared causal variants at the NACP and ACP risk loci

Finally, we extracted the prioritized shared causal variants, defined as the 95% credible set based on coloc's posterior probability for each variant to be causal for the shared signal ("SNP.PP.H4") at all loci with evidence of colocalization, defined by PP4 > 0.5 and PP4 / PP3 ≥ 5. The *CTRC* lead variant *rs497078* had the highest posterior probability to be the shared causal variant (SNP.PP.H4 = 0.92, **Table 2**) in NACP, followed by *rs545634* (SNP.PP.H4 = 0.08, **Table 2**), which were both associated with reduced *CTRC* expression ($Beta_{eQTL}$ = -0.31, P_{eQTL} = 1.80 x 10^{-10}, **Table 2**). In the ACP cohort the *CTRC* lead variant *rs545634* and the NACP lead variant *rs497078* had similar posterior probabilities to be the shared causal variant (SNP.PP.H4 = 0.51 and 0.49 respectively, **Table 2**). In ACP and NACP we discovered a single highly prioritized shared causal variant *rs12688220*, which was associated with increased *CLDN2* expression in the pancreas (SNP.PP.H4 = 1.00, $Beta_{eQTL}$ = 0.54, P_{eQTL} = 1.70 x 10^{-21}, **Table 2**).

Discussion

In this study, we present genome-wide significant GWAS hits at the *SPINK1* and *CTRC* loci in NACP. Furthermore, we report potential functional associations for both, NACP and ACP GWAS risk loci, from Bayesian colocalization analysis with the European pancreas eQTL public domain GTEx V8 dataset. Our analysis defines both, candidate causal genes, and shared causal variants.

Our findings of genome-wide significant association of common *SPINK1* and *CTRC* risk variants are in line with numerous former hypothesis driven approaches [4,27] and confirm previous results that effect sizes of *SPINK1* risk mutations were higher in NACP than ACP [4,28].

More than 90% of common variants show nominally significant eQTL associations across human tissues [29], implying a large number of false positive observations when comparing eQTL and GWAS data by simple overlaps. We applied Bayesian colocalization analysis [15] of both, NACP and ACP GWAS risk loci, and public domain pancreas eQTLs, to discover shared causal variants, candidate causal genes and disease causing mechanisms.

At the *CTRC* locus, we found evidence for a shared causal variant reducing *CTRC* expression, supporting a mechanism explaining the pathogenic potential of *CTRC* mutations. Our finding fits well in the established concept of trypsin-centred pathophysiology, as rare loss of function mutations in the gene impair the trypsin-degrading activity of chymotrypsin C and thus increase intra-pancreatic trypsin activity [3]. We discovered the two candidate shared causal variants *rs497078* and *rs545634* in ACP and NACP. Our findings in ACP are supported by a very recent GWAS-catalogue screen [30]. As both variants are in high LD, further functional studies are necessary to unravel the precise causal variant.

At the second NACP risk locus *SPINK1*, we found a number of risk variants in strong LD with the prominent p.N34S pancreatitis risk mutation [4]. We could not confirm previously suggested gene regulatory effects on *SPINK1* [31], as the European GTEx V8 pancreas dataset did not contain significant *SPINK1* eQTLs. Larger eQTL sample sizes in future GTEx releases or novel databases could help to elucidate causal mechanisms at this locus.

The first CP GWAS identified genome-wide significant risk variants at the *CLDN2-MORC4* locus, however, in subsequent functional analysis no significant correlation of the risk genotype with *CLDN2* gene expression levels was found [11]. Recently, however, *CLDN2-MORC4* risk variants were shown to be correlated with *CLDN2* expression levels in the pancreas [32]. Our colocalization analysis in ACP indicates the role of increased *CLDN2* expression in alcoholic chronic pancreatitis, probably driven by the shared causal variant *rs12688220*, again supported by a recent GWAS-catalogue screen [30]. Confirming this observation, we replicated the colocalization

and the shared causal variant *rs12688220* in NACP.

CLDN2 encodes a tight junction protein that controls the paracellular flow of molecules between epithelial cells. According to the Human Protein Atlas database single cell dataset, *CLDN2* expression is enhanced in pancreatic ductal, acinar and endocrine cells (https://www.proteinatlas.org/ENSG00000165376-CLDN2/single+cell+type/pancreas, Human Protein Atlas v21.0) [33,34]. Upregulation of *CLDN2* is associated with intestinal inflammation [35,36] and *CLDN2* deficiency in mice results in reduced cytokine production and T cell recruitment in experimental colitis [37]. Claudin-2 protein abundance has been proposed to be correlated with pancreatic tissue inflammation [11], supporting a role in inflammation also in the pancreas. Moreover, claudin-2 deficiency is associated with hypercalciuria in mice and kidney stone disease in humans [32], suggesting that claudin-2 may contribute to pancreatic calcium homeostasis. The recently identified chronic pancreatitis risk gene *TRPV6*, which codes for a calcium channel, also affects Ca^{2+} transport [10]. Thus, the risk variant dependent increased *CLDN2* expression may alter both, inflammatory processes and pancreatic calcium homeostasis. Experiments such as pancreas specific *CLDN2* knockout in mice are necessary to test this hypothesis.

At the *PRSS1-PRSS2* risk locus, previously suggested gene regulatory effects on *PRSS1* [11] could not be confirmed in this study, because the eQTL dataset did not contain significant *PRSS1* eQTLs. The high homology between trypsinogens could explain the absence of eQTLs, as read mapping in next generation sequencing could be corrupted. Instead, we obtained significant eQTLs for the homologous *PRSS2* overlapping with *PRSS1-PRSS2* ACP risk variants. However, we found no evidence for colocalization in ACP or NACP. Genotyping for GWAS and eQTL studies is typically performed with genomic DNA extracted from peripheral blood containing a considerable amount of T cells. This could lead to sequencing errors at the *PRSS1-PRSS2* locus, because it is located within the genomic VDJ-Region, encoding a set of T cell receptor genes undergoing genomic rearrangements (VDJ-recombination) and nucleotide additions in T cells. Thus, the genomic composition in this locus may differ considerably in blood and thereby corrupt the genotyping results in the entire genomic region including *PRSS1* and *PRSS2*.

Recently, we found ACP risk variants at the *CTRB1-CTRB2* locus in LD with a 16.6 kb inversion. The inversion exchanges the promoter regions of *CTRB1* and *CTRB2*, which modifies the expression ratio and of *CTRB1* and *CTRB2*, possibly leading to reduced anionic trypsinogen degradation [13]. Interestingly, *CTRB1* and *CTRB2* eQTLs did not colocalize with *CTRB1-CTRB2* risk variants. However, similar to the *PRSS1-PRSS2* locus, the *CTRB1-CTRB2* locus can be considered as a complex region [38], because of high homology of *CTRB1* and

CTRB2 and local structural variations [13,39], potentially leading to sequencing errors and corrupted colocalization analysis.

In summary, we found evidence for shared causal variants, which may modify CP-risk by altering pancreatic gene expression of *CTRC* at the *CTRC* locus and *CLDN2* at the *CLDN2-MORC4* locus. Although Bayesian colocalization can guide prioritization of causal genes and unravelling of molecular mechanisms at risk loci [15], further experimental approaches with our identified shared causal variants and genes are mandatory to confirm true causalities.

Acknowledgement

MEMBERS OF THE GESGA CONSORTIUM

Josef Frank, MSc,[1] Sven Cichon, PhD,[2,3] Jens Treutlein, PhD,[1] Monika Ridinger, MD,[4] Manuel Mattheisen, MD,[5,6,7,8] Per Hoffmann, PhD,[2,5,6] Stefan Herms, MSc,[2,5,6] Norbert Wodarz, MD,[4] Michael Soyka, MD,[9,10] Peter Zill, PhD,[10] Wolfgang Maier, MD,[11] Rainald Mössner, MD,[11] Wolfgang Gaebel, MD,[12] Norbert Dahmen, MD,[13] Norbert Scherbaum, MD,[14] Christine Schmäl, MD,[1] Michael Steffens, MD,[7] Susanne Lucae, MD,[15] Marcus Ising, MD,[16] Bertram Müller-Myhsok, MD,[17] Markus M Nöthen, MD,[5,6,18] Karl Mann, MD,[19] Falk Kiefer, MD,[19] and Marcella Rietschel, MD[1]

[1]Dept. of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, Germany
[2]Division of Medical Genetics, University Hospital Basel, Department of Biomedicine, University of Basel, Basel, Switzerland
[3]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Germany
[4]Dept. of Psychiatry, University Medical Center Regensburg, Univ. of Regensburg, Germany
[5]Dept. of Genomics, Life & Brain Center, Univ. of Bonn, Germany
[6]Inst. of Human Genetics, Univ. of Bonn, Germany
[7]Inst. for Medical Biometry, Informatics, and Epidemiology, Univ. of Bonn, Germany
[8]Department of Biomedicine, Aarhus University, Denmark
[9]Private Hospital Meiringen, Switzerland
[10]Dept. of Psychiatry, Univ. of Munich, Germany
[11]Dept. of Psychiatry, Univ. of Bonn, Germany
[12]Dept. of Psychiatry and Psychotherapy, Univ. of Düsseldorf, Germany
[13]Dept. of Psychiatry, Univ. of Mainz, Germany
[14]Addiction Research Group at the Department of Psychiatry and Psychotherapy, LVR Hospital Essen, University of Duisburg-Essen, Germany
[15]Dept. of Psychiatric Pharmacogenetics, Max-Planck-Institute of Psychiatry, Munich, Germany
[16]Dept. of Molecular Psychology, Max-Planck-Institute of Psychiatry, Munich, Germany
[17]Dept. of Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany
[18]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
[19]Dept. of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, Germany

References

[1] Mayerle J, Sendler M, Hegyi E, Beyer G, Lerch MM, Sahin-Tóth M. Genetics, Cell Biology, and Pathophysiology of Pancreatitis. Gastroenterology 2019;156:1951-1968.e1. https://doi.org/10.1053/j.gastro.2018.11.081.

[2] Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, et al. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. Nat Genet 1996;14:141–5. https://doi.org/10.1038/ng1096-141.

[3] Sahin-Tóth M, Tóth M. Gain-of-Function Mutations Associated with Hereditary Pancreatitis Enhance Autoactivation of Human Cationic Trypsinogen. Biochemical and Biophysical Research Communications 2000;278:286–9. https://doi.org/10.1006/bbrc.2000.3797.

[4] Witt H, Luck W, Hennies HC, Claßen M, Kage A, Laß U, 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. https://doi.org/10.1038/76088.

[5] Rosendahl J, Witt H, Szmola R, Bhatia E, Ózsvári B, Landt O, et al. Chymotrypsin C (CTRC) variants that diminish activity or secretion are associated with chronic pancreatitis. Nat Genet 2008;40:78–82. https://doi.org/10.1038/ng.2007.44.

[6] Lasher D, Szabó A, Masamune A, Chen J-M, Xiao X, Whitcomb DC, et al. Protease-Sensitive Pancreatic Lipase Variants Are Associated With Early Onset Chronic Pancreatitis. Am J Gastroenterol 2019;114:974–83. https://doi.org/10.14309/ajg.0000000000000051.

[7] Fjeld K, Weiss FU, Lasher D, Rosendahl J, Chen J-M, Johansson BB, et al. A recombined allele of the lipase gene CEL and its pseudogene CELP confers susceptibility to chronic pancreatitis. Nat Genet 2015;47:518–22. https://doi.org/10.1038/ng.3249.

[8] Witt H, Beer S, Rosendahl J, Chen J-M, Chandak GR, Masamune A, et al. Variants in CPA1 are strongly associated with early onset chronic pancreatitis. Nat Genet 2013;45:1216–20. https://doi.org/10.1038/ng.2730.

[9] Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, et al. Mutations of the Cystic Fibrosis Gene in Patients with Chronic Pancreatitis. N Engl J Med 1998;339:645–52. https://doi.org/10.1056/NEJM199809033391001.

[10] Masamune A, Kotani H, Sörgel FL, Chen J-M, Hamada S, Sakaguchi R, et al. Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis. Gastroenterology 2020;158:1626-1641.e8. https://doi.org/10.1053/j.gastro.2020.01.005.

[11] Alzheimer's Disease Genetics Consortium, Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, 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. https://doi.org/10.1038/ng.2466.

[12] Giri AK, Midha S, Banerjee P, Agrawal A, Mehdi SJ, Dhingra R, et al. Common Variants in CLDN2 and MORC4 Genes Confer Disease Susceptibility in Patients with Chronic Pancreatitis. PLoS ONE 2016;11:e0147345. https://doi.org/10.1371/journal.pone.0147345.

[13] Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, et al. Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut 2018;67:1855–63. https://doi.org/10.1136/gutjnl-2017-314454.

[14] Derikx MH, Kovacs P, Scholz M, Masson E, Chen J-M, Ruffert C, 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. https://doi.org/10.1136/gutjnl-2014-307453.

[15] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet 2014;10:e1004383. https://doi.org/10.1371/journal.pgen.1004383.

[16] The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020;369:1318–30. https://doi.org/10.1126/science.aaz1776.

[17] Löhr JM, Dominguez- Munoz E, Rosendahl J, Besselink M, Mayerle J, Lerch MM, et al. United European Gastroenterology evidence- based guidelines for the diagnosis and therapy of chronic pancreatitis (HaPanEU). United European Gastroenterol j 2017;5:153–99. https://doi.org/10.1177/2050640616684695.

[18] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 2015;4:7. https://doi.org/10.1186/s13742-015-0047-8.

[19] de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Human Molecular Genetics 2008;17:R122–8. https://doi.org/10.1093/hmg/ddn288.

[20] Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Bioinformatics; 2014. https://doi.org/10.1101/005165.

[21] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65. https://doi.org/10.1038/nature11632.

[22] Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics 2009;25:1841–2. https://doi.org/10.1093/bioinformatics/btp328.

[23] Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 2009;4:1184–91. https://doi.org/10.1038/nprot.2009.97.

[24] Schmiedel BJ, Rocha J, Gonzalez-Colin C, Bhattacharyya S, Madrigal A, Ottensmeier CH, et al. COVID-19 genetic risk variants are associated with expression of multiple genes in diverse immune cell types. Nat Commun 2021;12:6760. https://doi.org/10.1038/s41467-021-26888-3.

[25] Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer; 2016. https://doi.org/10.1007/978-3-319-24277-4.

[26] Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. Nat Genet 2019;51:768–9. https://doi.org/10.1038/s41588-019-0404-0.

[27] LaRusch J, Lozano-Leon A, Stello K, Moore A, Muddana V, O'Connell M, 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. Clinical and Translational Gastroenterology 2015;6:e68. https://doi.org/10.1038/ctg.2014.13.

[28] Witt H. Mutation in the SPINK1 Trypsin Inhibitor Gene, Alcohol Use, and Chronic Pancreatitis. JAMA 2001;285:2716. https://doi.org/10.1001/jama.285.21.2716-a.

[29] GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 2017;550:204–13. https://doi.org/10.1038/nature24277.

[30] King EA, Dunbar F, Davis JW, Degner JF. Estimating colocalization probability from limited summary statistics. BMC Bioinformatics 2021;22:254. https://doi.org/10.1186/s12859-021-04170-z.

[31] Boulling A, Masson E, Zou W, Paliwal S, Wu H, Issarapu P, et al. Identification of a functional enhancer variant within the chronic pancreatitis- associated *SPINK1* c.101A>G (p.Asn34Ser)- containing haplotype. Human Mutation 2017;38:1014–24. https://doi.org/10.1002/humu.23269.

[32] Curry JN, Saurette M, Askari M, Pei L, Filla MB, Beggs MR, et al. Claudin-2 deficiency associates with hypercalciuria in mice and human kidney stone disease. Journal of Clinical Investigation 2020;130:1948–60. https://doi.org/10.1172/JCI127750.

[33] Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science 2015;347:1260419–1260419. https://doi.org/10.1126/science.1260419.

[34] Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single–cell type transcriptomics map of human tissues. Sci Adv 2021;7:eabh2169. https://doi.org/10.1126/sciadv.abh2169.

[35] Weber CR, Nalle SC, Tretiakova M, Rubin DT, Turner JR. Claudin-1 and claudin-2 expression is elevated in inflammatory bowel disease and may contribute to early neoplastic transformation. Lab Invest 2008;88:1110–20. https://doi.org/10.1038/labinvest.2008.78.

[36] Luettig J, Rosenthal R, Barmeyer C, Schulzke J. Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation. Tissue Barriers 2015;3:e977176. https://doi.org/10.4161/21688370.2014.977176.

[37] Raju P, Shashikanth N, Tsai P-Y, Pongkorpsakol P, Chanez-Paredes S, Steinhagen PR, et al. Inactivation of paracellular cation-selective claudin-2 channels attenuates immune-mediated experimental colitis in mice. Journal of Clinical Investigation 2020;130:5197–208. https://doi.org/10.1172/JCI138697.

[38] Seltsam K, Pentner C, Weigl F, Sutedjo S, Zimmer C, Beer S, et al. Sequencing of the complex CTRB1-CTRB2 locus in chronic pancreatitis. Pancreatology 2020;20:1598–603. https://doi.org/10.1016/j.pan.2020.09.017.

[39] Jermusyk A, Zhong J, Connelly KE, Gordon N, Perera S, Abdolalizadeh E, et al. A 584 bp deletion in CTRB2 inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer. The American Journal of Human Genetics 2021;108:1852–65. https://doi.org/10.1016/j.ajhg.2021.09.002.

Figure 1 Manhattan plot of the NACP-GWAS dataset. The red line indicates genome-wide significance after Bonferroni correction (5×10^{-8}).

Figure 2 Visualization of colocalization in ACP. Scatter plots for the indicated eQTL genes at the CP risk loci *CLDN2-MORC4* (A), *CTRB1-CTRB2* (B,C,E,F), *CTRC* (D) and *PRSS1-PRSS2* (G,H). Selected (clusters of) top variants biased towards eQTL or GWAS signal strength were manually highlighted in red and green, respectively.

Figure 3 Visualization of colocalization in NACP. Scatter plots for the indicated eQTL genes at the CP risk loci *CLDN2-MORC4* (A), *CTRB1-CTRB2* (B,C,E,F), *CTRC* (D) and *PRSS1-PRSS2* (G,H). Selected (clusters of) top variants biased towards eQTL or GWAS signal strength were manually highlighted in red and green, respectively.

Table 1 Colocalization of GWAS risk loci with pancreas eQTLs

Disease	Locus	Gene	PP3	PP4	PP4/PP3
ACP	*CLDN2-MORC4*	*CLDN2*	0.02	0.98	42.20
	CTRB1-CTRB2	*CTRB1*	0.86	0.13	0.16
	CTRB1-CTRB2	*CTRB2*	0.96	0.04	0.04
	CTRC	*CTRC*	0.01	0.99	95.51
	CTRB1-CTRB2	ENSG00000240338	0.30	0.70	2.31
	CTRB1-CTRB2	ENSG00000280152	0.93	0.07	0.07
	PRSS1-PRSS2	*PRSS2*	0.25	0.64	2.57
	PRSS1-PRSS2	*TRBV29-1*	1.00	0.00	0.00
NACP	*CLDN2-MORC4*	*CLDN2*	0.09	0.67	7.18
	CTRB1-CTRB2	*CTRB1*	0.25	0.49	1.93
	CTRB1-CTRB2	*CTRB2*	0.25	0.49	1.94
	CTRC	*CTRC*	0.01	0.99	95.46
	CTRB1-CTRB2	ENSG00000240338	0.48	0.04	0.09
	CTRB1-CTRB2	ENSG00000280152	0.34	0.33	0.99
	PRSS1-PRSS2	*PRSS2*	0.39	0.36	0.91
	PRSS1-PRSS2	*TRBV29-1*	0.37	0.58	1.59

Evidence of colocalization was defined by PP4 > 0.5 and PP4/PP3 ≥ 5. PP3, posterior probability for association with CP risk and eQTLs, but independent causal variants; PP4, posterior probability for association with CP risk and eQTLs with a single shared causal variant.

Table 2 Candidate shared causal variants at GWAS risk loci

Disease	Locus	Candidate causal gene	Candidate causal variant	Counted allele	P_{GWAS}	$Beta_{GWAS}$	P_{eQTL}	$Beta_{eQTL}$	SNP.PP.H4
ACP	*CTRC*	*CTRC*	*rs545634*	A	2.16×10^{-22}	0.61	1.80×10^{-10}	-0.31	0.51
	CTRC	*CTRC*	*rs497078*	T	2.47×10^{-22}	0.60	1.80×10^{-10}	-0.31	0.49
	CLDN2-MORC4	*CLDN2*	*rs12688220*	T	9.54×10^{-33}	0.45	1.70×10^{-21}	0.54	1.00
NACP	*CTRC*	*CTRC*	*rs497078*	T	1.22×10^{-21}	0.83	1.80×10^{-10}	-0.31	0.92
	CTRC	*CTRC*	*rs545634*	A	1.97×10^{-20}	0.81	1.80×10^{-10}	-0.31	0.08
	CLDN2-MORC4	*CLDN2*	*rs12688220*	T	1.08×10^{-4}	0.23	1.70×10^{-21}	0.54	1.00

The 95% credible sets of shared causal variants between both, the NACP and the ACP GWAS risk loci, and pancreas eQTLs from GTEx were extracted from coloc result objects for all loci with evidence of colocalization, defined by PP4 > 0.5 and PP4/PP3 ≥ 5. SNP.PP.H4, posterior probability of the variant to be causal for the shared signal.

Figure 1

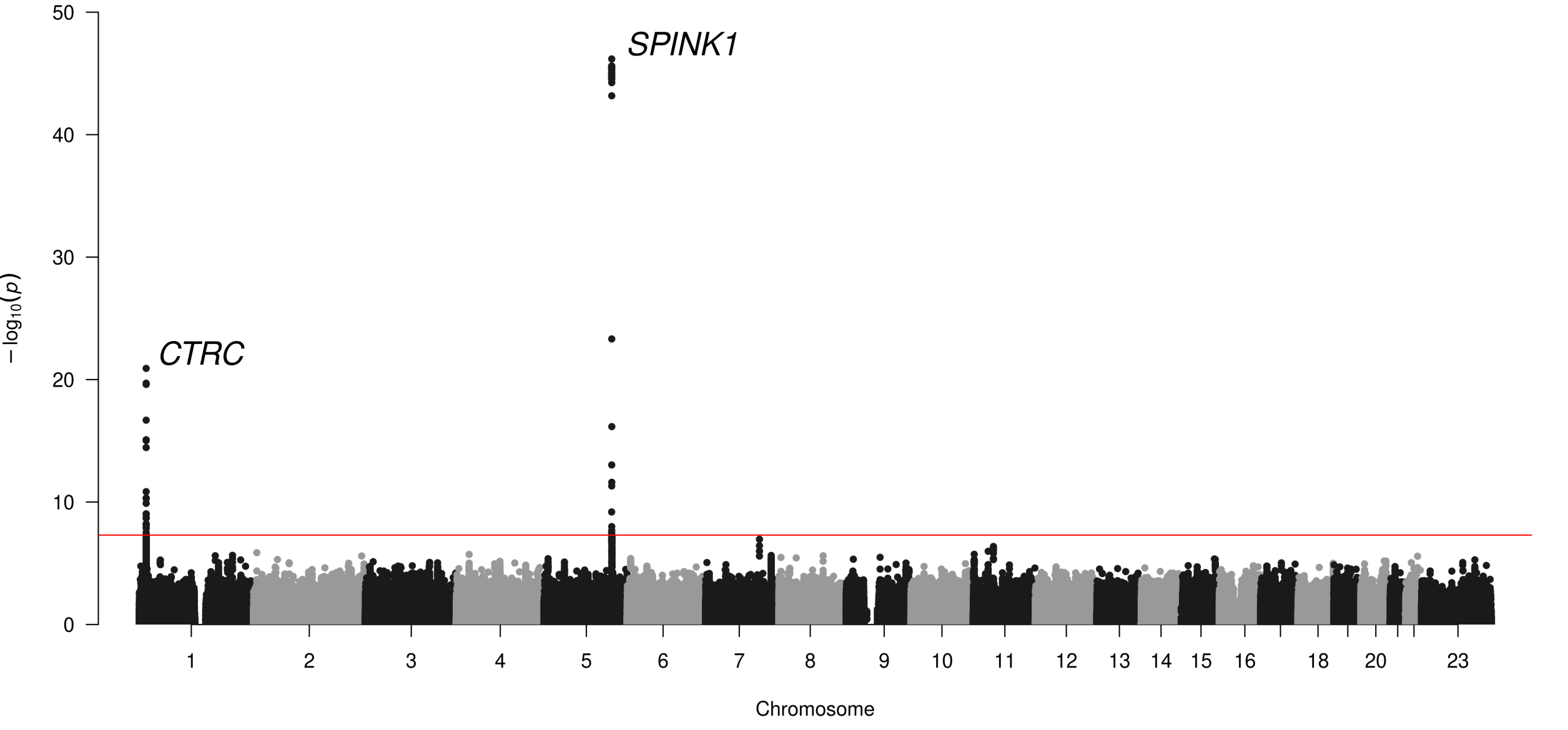


Figure 2

Click here to access/download;Figure;Figure_2.pdf

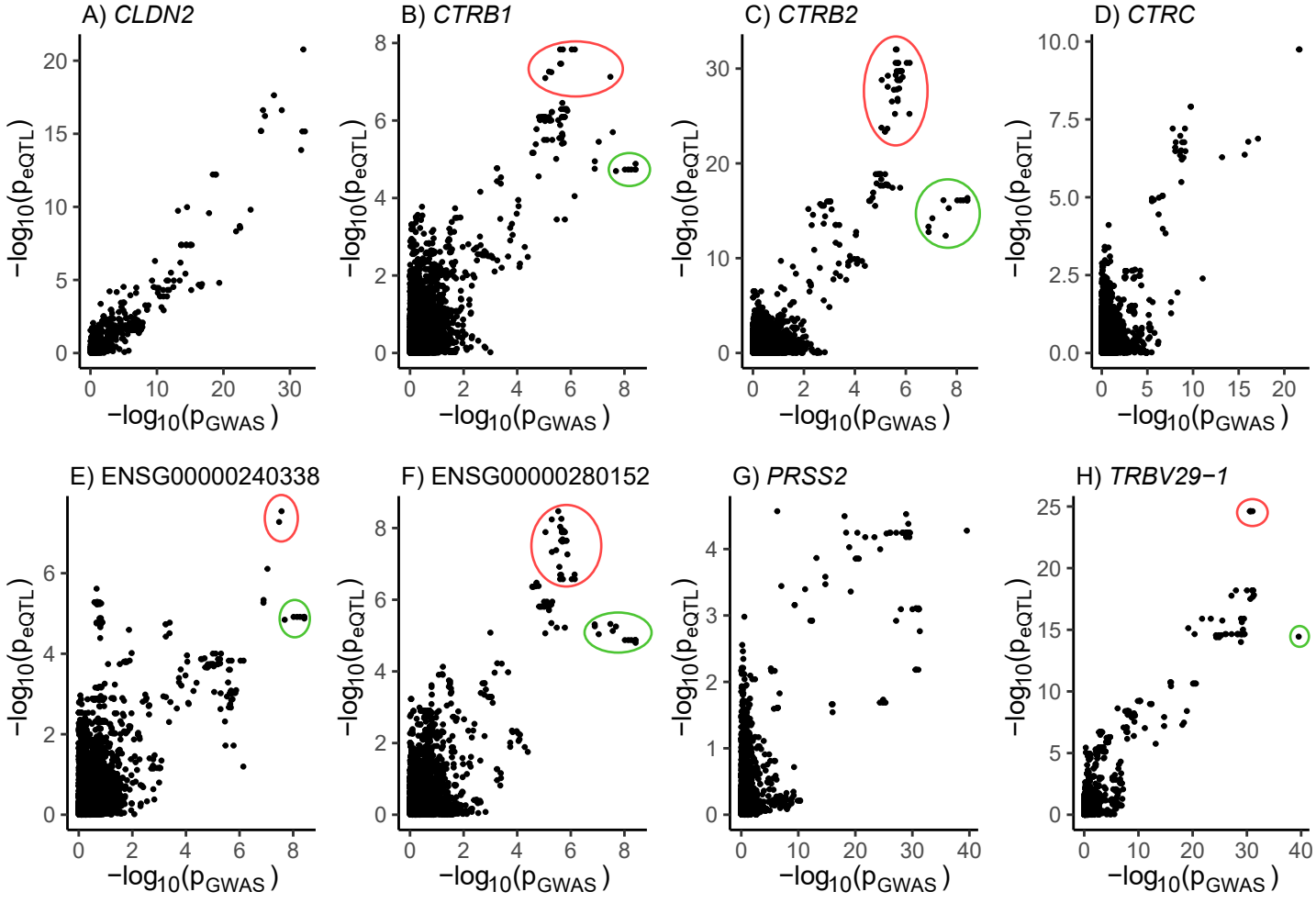
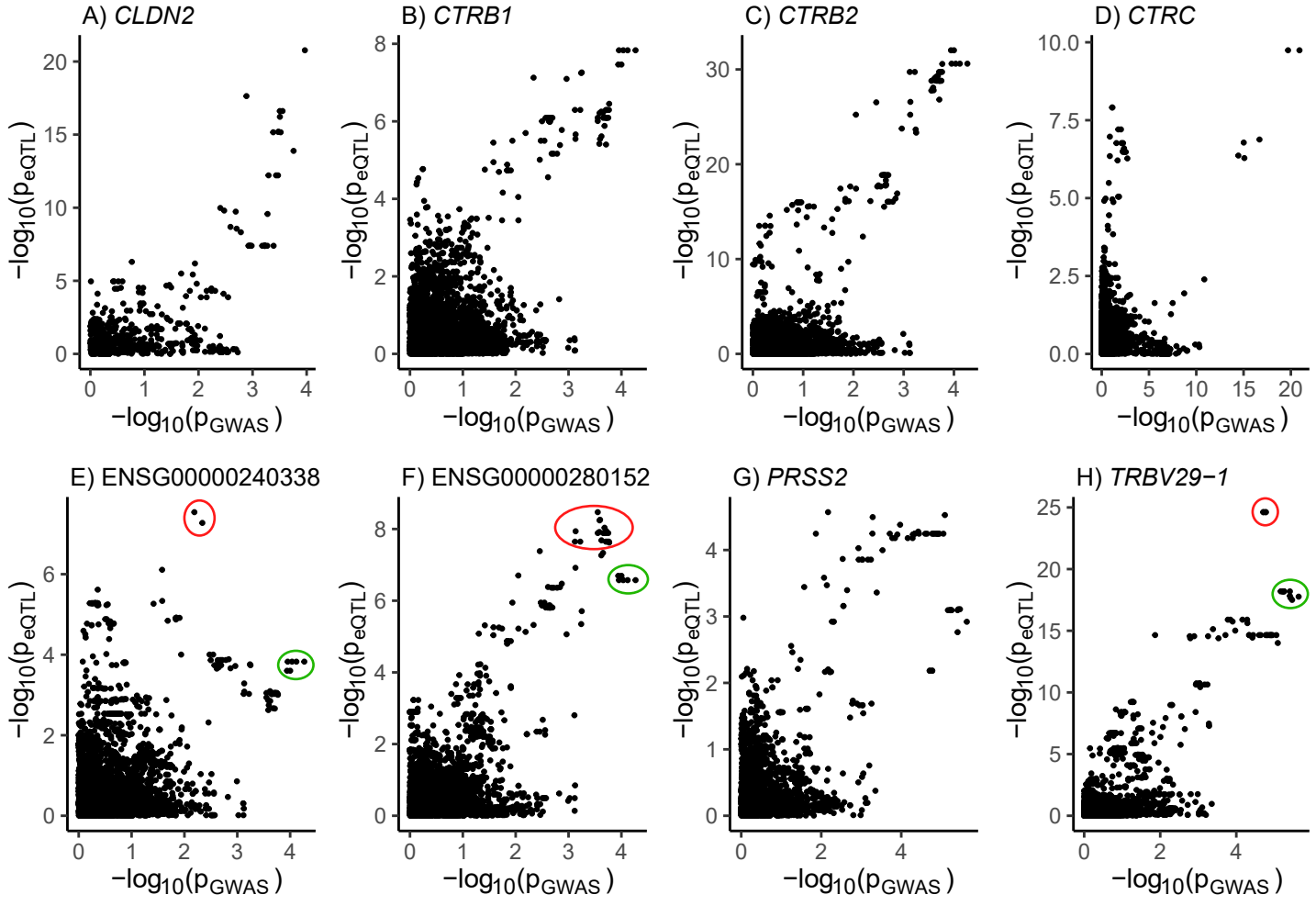


Figure 3

Click here to access/download;Figure;Figure_3.pdf



Supplementary Interactive Plot Data (CSV)

Click here to access/download

Supplementary Interactive Plot Data (CSV)

Supplementary_tables.xlsx